ISSN 2073-7394

Національний університет
“Полтавська політехніка імені Юрія Кондратюка”

National University
“Yuri Kondratyuk Poltava Polytechnic”

Системи управління, навігації та зв'язку

Control, navigation and communication systems

Випуск 1 (75) **Issue 1 (75)**

Щоквартальне видання

Засноване у 2007 році

У журналі відображені результати наукових досліджень з розробки та удосконалення систем управління, навігації та зв’язку у різних проблемних галузях.

Засновник і видавець:
Національний університет
“Полтавська політехніка імені Юрія Кондратюка”

Телефон:
+38 (050) 302-20-71

E-mail редколегії:
kuchuk_nina@ukr.net

Інформаційний сайт:
http://journals.nupp.edu.ua/sunz

Quarterly

Founded in 2007

Journal represent the research results on the development and improvement of control, navigation and communication systems in various areas

Founder and publisher:
National University
"Yuri Kondratyuk Poltava Polytechnic"

Phone:
+38 (050) 302-20-71

E-mail of the editorial board:
kuchuk_nina@ukr.net

Information site:
http://journals.nupp.edu.ua/sunz

За достовірність викладених фактів, цитат та інших відомостей відповідальність несе автор

*Журнал індексується міжнародними наукометричними базами: Index Copernicus (ICV = **82.05**), General Impact Factor, Google Scholar, Academic Resource Index, Scientific Indexed Service*

*Затверджений до друку Вченою Радою Національного університету “*Полтавська політехніка імені Юрія Кондратюка*” (протокол від 09 лютого 2024 року № 2).*

Свідоцтво про державну реєстрацію КВ № 24464-14404 ПР від 27.03.2020 р.

Включений до “Переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук, кандидата наук та ступеня доктора філософії ” до категорії Б – наказами МОН України від 17.03.2020 № 409 та від 09.02.2021 № 157

Полтава • 2024

З М І С Т

АВТОМОБІЛЬНИЙ, РІЧКОВИЙ, МОРСЬКИЙ ТА АВІАЦІЙНИЙ ТРАНСПОРТ

УПРАВЛІННЯ В СКЛАДНИХ СИСТЕМАХ

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ЦИВІЛЬНА БЕЗПЕКА

ЗВ'ЯЗОК, ТЕЛЕКОМУНІКАЦІЇ ТА РАДІОТЕХНІКА

Організації авторів

Азербайджанський технічний університет, Баку, Азербайджан
Національний університет оборони, Баку, Азербайджан
Державний НДІ випробувань і сертифікації озброєння та військової техніки, Черкаси, Україна
Київський національний університет будівництва і архітектури, Київ, Україна
Кременчуцький національний університет імені Михайла Остроградського, Кременчук, Україна
Льотна академія Національного авіаційного університету, Кропивницький, Україна
Морський фаховий коледж Херсонської державної морської академії, Херсон, Україна
Науково-дослідний, проектно-конструкторський та технологічний інститут мікрографії, Харків, Україна
Національний авіаційний університет, Київ, Україна
Національний аерокосмічний університет імені М. Є. Жуковського «ХАІ», Харків, Україна
Національний технічний університет «Харківський політехнічний інститут», Харків, Україна
Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ, Україна
Національний транспортний університет, Київ, Україна
Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава. Україна
Національний університет цивільного захисту України, Харків, Україна
Одеський національний морський університет, Україна
Харківський національний університет радіоелектроніки, Харків, Україна
Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна
Український державний університет залізничного транспорту, Харків, Україна

Authors affiliation

Azerbaijan Technical University, Baku, Azerbaijan
National Defense University, Baku, Azerbaijan
SR Institute of Testing and Certification of Weapons and Military Equipment, Cherkasy, Ukraine
Kyiv National University of Construction and Architecture, Kyiv, Ukraine
Kremenchuk Mykhailo Ostrohradskyi National University, Kremenchuk, Ukraine
Flight Academy of the National Aviation University, Kropyvnytskyi, Ukraine
Maritime College of the Kherson State Maritime Academy, Kherson, Ukraine
Research, Design and Technology Institute of Micrography, Kharkiv, Ukraine
National Aviation University, Kyiv, Ukraine
National Aerospace University named after M.E. Zhukovsky "KHAI", Kharkiv, Ukraine
National Technical University “Kharkiv Polytechnic Institute”, Kharkiv, Ukraine
National Technical University “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine
National Transport University, Kyiv, Ukraine
National University “Yuri Kondratyuk Poltava Polytechnic”, Poltava, Ukraine
National University of Civil Defense of Ukraine, Kharkiv, Ukraine
Odessa National Maritime University, Ukraine
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine
Ukrainian State University of Railway Transport, Kharkiv, Ukraine

Автомобільний, річковий, морський та авіаційний транспорт

УДК 629.5.078:656.075 **doi: 10.26906/SUNZ.2024.1.005**

А. І. Головань

Одеський національний морський університет, Одеса, Україна

ОСОБЛИВОСТІ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ СИСТЕМ ТЕХНІЧНОГО ОБСЛУГОВУВАННЯ ВАНТАЖНИХ СУДЕН

Анотація. У статті розглянуто важливу проблему оцінювання ефективності систем технічного обслуговування вантажних суден. Аналіз останніх досліджень і публікації в цій області, вказує на необхідність оцінювання ефективності технічного обслуговування вантажних суден. Метою статті є розроблення і формування алгоритмів інформаційної системи моніторингу показників ефективності технічного обслуговування вантажних суден для застосування їх в інформаційній системи оцінювання способів підвищення ефективності технічного обслуговування. Результати статті включають блок-схеми алгоритмів: збору даних про параметри системи технічного обслуговування та технічного стану суднових технічних засобів і конструкцій, моніторингу і визначення статусу несправностей та оцінювання способів підвищення ефективності системи технічного обслуговування з можливістю прогнозування показників ефективності системи технічного обслуговування. Висновки підкреслюють важливість аналізу експлуатаційних даних, використання інформаційно-комунікаційних технологій та системного підходу до оцінювання способів підвищення ефективності з метою забезпечення надійності і ефективності технічного обслуговування вантажних суден.

Ключові слова: технічне обслуговування, вантажні судна, ефективність, прогнозування, аналіз.

Вступ

Постановка проблеми. Системи моніторингу технічного стану суднових технічних засобів, конструкцій (СТЗіК) та експлуатаційних характеристик системи технічного обслуговування (ТО) в умовах Інтелектуальних транспортних систем розроблені для забезпечення постійного автоматизованого контролю за технічними параметрами суднових технічних засобів, конструкцій, систем і комплексів вантажних суден. Вони призначені для виявлення відмовних станів і запобігання їх подальшому розвитку. Крім того, ці системи дозволяють налагоджувати системи комплексного прескриптивного технічного обслуговування з можливістю прогнозування технічного стану та обґрунтованого забезпечення оптимальних параметрів системи ТО.

Зазвичай такі системи представляють собою складний комплекс бортових і стаціонарних технічних і програмних засобів. Розробка системи моніторингу технічного стану в автономному виконанні та системи моніторингу експлуатаційних характеристик системи ТО вимагає значних інтелектуальних, часових і матеріальних ресурсів. Кожне окреме вантажне судно, обладнане такою системою, потребує інвестувати фінансові ресурси в установку не лише бортових діагностичних комплексів, але й у пристрої обробки інформації, системи зв'язку та сигналізації для виявлення відмовних станів.

Головною проблемою, яка висвітлюється в статті, є необхідність розроблення алгоритмів і впровадження їх у єдину систему моніторингу вантажних суден з метою забезпечення постійного контролю за різними показниками ефективності системи технічного обслуговування, підвищення надійності та безпеки судноплавства.

Аналіз останніх досліджень і публікацій. В цілому, аналіз останніх досліджень і публікацій вказує на важливість впровадження інформаційних систем моніторингу ефективності технічного обслуговування суден. Важливими постають питання: системного технічного обслуговування, що орієнтоване на надійність (SRCM), яке ефективно підвищує надійність і безпеку повністю електричних суден шляхом визначення оптимальних завдань технічного обслуговування і використання методів управління ризиками, таких як аналіз рівнів захисту (LOPA) [1]; оцінювання ефективності технічного обслуговування суден за допомогою підходу до аналізу процесів, заснованого на вхідних, контрольних, вихідних даних і ресурсах (ICOR) [2]; оцінювання ефективності технічного обслуговування суден шляхом застосування моделі на основі Марківського аналізу для оцінки надійності кожної підсистеми, а потім розгляд шляхів підвищення надійності та вимірювання економічних переваг [3]; автоматизації оцінки стану танків і конструкцій вантажних суден за допомогою бездротової технології, яка значно скорочує час і зусилля, максимізуючи ефективність, економічність і надійність суден [4]. Метод оцінки життєвого циклу збільшує значення операційного показника енергоефективності (EEOI) в реальному часі, особливо при менших обертах двигуна і меншій осадці судна, оцінюючи ефективність технічного обслуговування судна [5]. Періодичне вимірювання швидкості/потужності може дозволити екіпажу судна спрогнозувати відповідний час для технічного обслуговування корпусу, мінімізуючи споживання палива та зменшуючи забруднення навколишнього середовища [6]. Запропонована стратегія мінімізує витрати на паливо, викиди та технічне обслуговування повністю електричних суден шляхом оптимального планування рейсів, періодів генерації

та оцінки ризиків з урахуванням інформації про моніторинг технічного стану [7]. Автоматизована служба моніторингу ефективності технічних операцій покращує і підтримує ефективність виробництва електроенергії на судні шляхом моніторингу роботи суднових головних і допоміжних дизельних двигунів [8]. Аналіз даних може допомогти оцінити надійність суднових двигунів і спланувати заходи з технічного обслуговування, що потенційно зменшить кількість відмов двигунів і аварій в судноплавній галузі [9]. Модель коригування технічного обслуговування складається з аналізу експлуатаційних даних та аналізу ризиків, визначення важливих компонентів та прийняття ефективної політики технічного обслуговування [10].

Всі ці дослідження разом підкреслюють важливість аналізу експлуатаційних даних, використання інформаційно-комунікаційних технологій і розгляду різних підходів до вимірювання ефективності технічного обслуговування вантажних суден.

Мета статті полягає в розробці і формуванні алгоритмів інформаційної системи моніторингу показників ефективності технічного обслуговування вантажних суден.

Основний матеріал

Науковою та технічною оптимальністю для здійснення інтелектуального моніторингу технічного стану СТЗіК, а також експлуатаційних характеристик системи ТО є інтегрована система. Ця система включає в себе поєднання стандартного та додаткового обладнання для інформаційно-діагностичного моніторингу, яке програмно вбудовано в навігаційно-зв'язковий комплекс судна та виконує функції, пов'язані з супутниковою навігацією.

З метою виконання моніторингу експлуатаційних характеристик системи ТО, визначення технічного стану та ідентифікації несправностей в СТЗіК та системах вантажних суден необхідно об'єднати навігаційно-зв'язкові та діагностичні компоненти. Це об'єднання передбачає технологічний і програмний зв'язок між ними та створення розгалуженої системи контролю експлуатаційних характеристик системи ТО, а також робочих параметрів окремих СТЗіК, систем і комплексів вантажних суден. При цьому інтеграція бортового навігаційного обладнання з основними технологічними компонентами системи моніторингу технічного стану СТЗіК повинна відбуватися в рамках єдиної концепції мобільної інформаційно-діагностичної системи.

Необхідність створення комплексної автоматизованої системи в контексті інформаційних умов Інтелектуальної транспортної системи випливає з необхідності вирішення низки складних завдань, які взаємодіють між собою та обмежені апаратно-програмними можливостями конкретної мікропроцесорної техніки СТЗіК, систем і комплексів.

Процеси керування сучасних суднових технічних засобів, систем і комплексів базуються на використанні мікроконтролерів, які володіють розширеним набором засобів зв'язку. Це дозволяє здійснювати збір інформації від стандартних датчиків суднових технічних засобів, систем і комплексів, проводити часткову обробку вимірюваних даних, формувати діагностичні повідомлення та передавати інформацію через діагностичні інтерфейси On-Board Diagnostics - Marine (OBD-M). Аналіз технологічних рішень, що використовуються на ринку обслуговування вантажних суден, виявив відсутність можливості забезпечення повноцінного аналізу не лише отриманих характеристик системи ТО і параметрів технічного стану СТЗіК, систем і комплексів, а також прогнозування їхнього стану.

Актуальність проблеми прогнозування технічного стану та ефективності системи ТО вантажних суден у контексті сучасних вимог до систем керування судновими технічними засобами, системами та комплексами, які підлягають комплексному прескриптивному технічному обслуговуванню, беззаперечна. Важливо не лише перевіряти справність СТЗіК та ефективність системи ТО на поточний момент (під час контролю), але й забезпечувати їх тривалу працездатність та постійне підвищення ефективності системи технічного обслуговування вантажних суден протягом передбачуваного часового інтервалу.

1. **Формування процесів збору даних про показники ефективності системи ТО та параметри технічного стану СТЗіК.** Алгоритм збору даних (рис. 1) про параметри системи технічного обслуговування та технічного стану суднових технічних засобів і конструкцій в межах комплексної системи прескриптивного технічного обслуговування починається з ініціалізації збору даних. Цей крок включає запуск системи збору даних та встановлення необхідних параметрів, таких як часові інтервали, типи датчиків та області моніторингу. Далі відбувається автоматичний збір даних з різних датчиків на судні, таких як температура, тиск, вібрації, рівні зносу тощо. Агрегація даних виконується у вигляді розрахунку середнього значення:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}, \tag{1}$$

де $\bar{X}$ – середнє значення, x_i – значення вимірювання, n – кількість вимірювань.

Після збору даних відбувається їх первинна обробка та валідація, яка включає перевірку на наявність помилок або відхилень та виключення аномальних даних. Аналіз даних проводиться за допомогою статистичного аналізу для визначення тенденцій та закономірностей.

Важливими інструментами в цьому процесі є формули для визначення стандартного відхилення або варіабельності:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2}, \tag{2}$$

де s – стандартне відхилення, x_i – значення вимірювання, $\bar{X}$ – середнє значення, n - кількість вимірювань.

На наступному етапі виконується класифікація інформації про стан системи технічного обслуговування і технічний стан суднових технічних засобів.

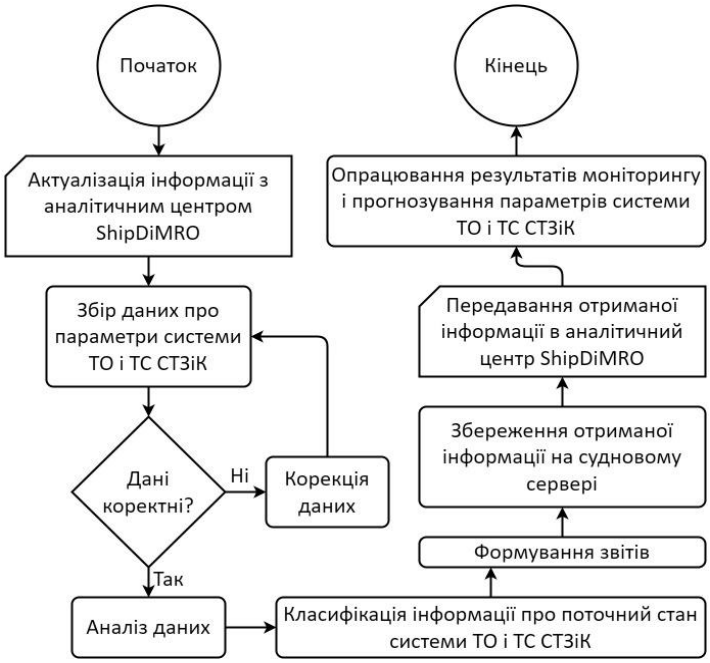


Рис. 1. Блок-схема алгоритму збору даних про параметри системи технічного обслуговування та технічного стану суднових технічних засобів і конструкцій в межах комплексної системи прескриптивного технічного обслуговування

Формуються звіти на основі зібраних даних, які подаються у зрозумілому та структурованому форматі для подальшого аналізу. Дані зберігаються на судновому сервері і передаються для інтеграції з аналітичним центром системи технічного обслуговування ShipDiMRO, де вони використовуються для планування технічного обслуговування та ремонтних робіт. На завершальному етапі прогнозуються параметри системи технічного обслуговування. Представлений алгоритм забезпечує систематичний підхід до збору та аналізу даних, що є важливим для підтримання ефективності системи технічного обслуговування та забезпечення технічної безпеки судна.

Алгоритм збору даних, представлений на Рис. 1, є ключовим компонентом інформаційної моделі системи комплексного прескриптивного технічного обслуговування «ShipDiMRO» і спрямований на збір даних про параметри системи технічного обслуговування і параметри суднових технічних засобів і конструкцій. Цей процес відіграє вирішальну роль у забезпеченні точності прогнозування параметрів системи технічного обслуговування і параметрів технічного стану суднових технічних засобів і конструкцій. Одним із найважливіших аспектів процесу є вибір часового інтервалу для отримання інформації. Зменшення цього інтервалу забезпечує вищу точність прогнозу, але одночасно збільшує час необхідний для обчислення прогнозних значень.

Необхідними для прогнозування є дані, які представляють собою послідовності впорядкованих в часі числових показників основних параметрів системи технічного обслуговування і параметрів технічного стану суднових технічних засобів і конструкцій. Ці дані формують повні інтервальні часові ряди, що є критично важливими для точного прогнозування. Паралельно зі збором даних про параметри системи технічного обслуговування і технічного стану суднових технічних засобів і конструкцій проводиться моніторинг та визначення їх потенційних несправностей.

Розроблені спеціалізовані алгоритми збору даних та розпізнавання стану несправностей, які є невід'ємною частиною процесу прогнозування параметрів системи технічного обслуговування. Ці алгоритми адаптовано до специфічних умов використання в межах віртуального підприємства «shipmonitoring.org». Визначним для алгоритму є інтервал часу Δt для зчитування інформації з датчиків та загальний період T, протягом якого відбувається збір інформації. В результаті, вихідними даними алгоритму є масив даних, який містить часові ряди значень параметрів. Ключові етапи та особливості процесу збору даних детально представлені на згаданому рис. 1.

2. **Особливості процесу діагностування ефективності системи ТО і визначення статусу несправностей СТЗ і К у складі системи моніторингу комплексної системи прескриптивного технічного обслуговування.** Процес моніторингу (рис. 2) та визначення статусу несправностей системи технічного обслуговування вантажного судна, згідно з розробленим алгоритмом, розпочинається з актуалізації інформації з аналітичним центром ShipDiMRO. Це забезпечує актуальність даних у системі. Далі відбувається ініціалізація системи технічного обслуговування, що включає активацію необхідних модулів для моніторингу та управління обладнанням. Після цього система проходить через етап самодіаг-ностики, який визначає її готовність і функціональний стан.

У разі виявлення зниження рівня технічної придатності судна, стандартне відхилення якого розраховується за формулою (3), система аналізує показники ефективності системи технічного обслуговування:

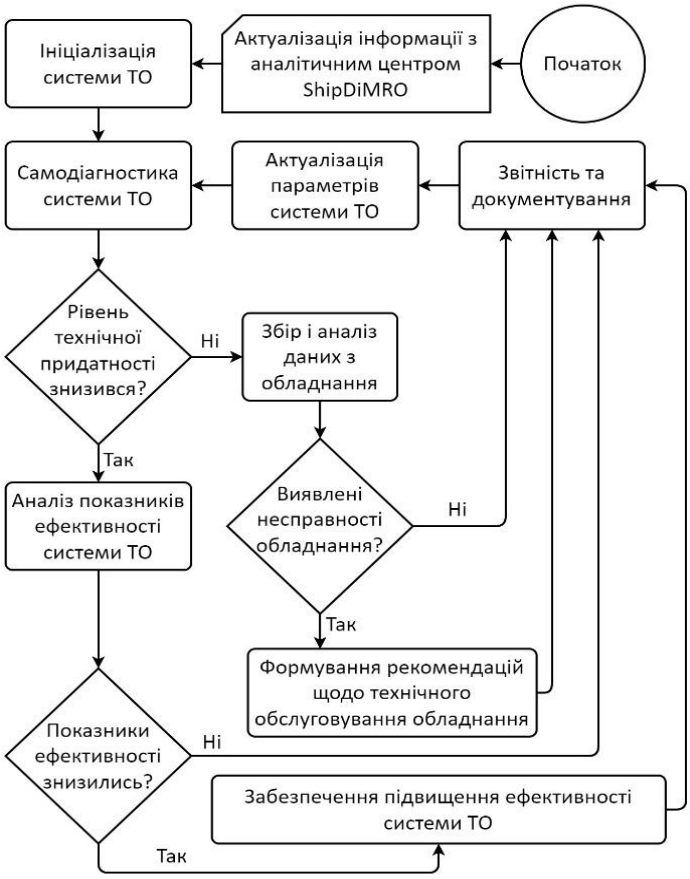


Рис. 2. Блок-схема алгоритму моніторингу і визначення статусу несправностей системи технічного обслуговування вантажного судна в межах комплексної системи прескриптивного технічного обслуговування

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}, \quad (3)$$

де σ – стандартне відхилення, x_i – окремі вимірювання, μ – середнє значення, N – кількість вимірювань.

Якщо ж технічна придатність не знизилася, здійснюється збір та аналіз даних безпосередньо з обладнання судна. Для цього використовується статистичний аналіз даних, наприклад, обчислення середніх значень, аналіз трендів або лінійна регресія (4):

$$y = a \cdot x + b, \quad (4)$$

де y – залежна змінна, x – незалежна змінна, a, b – коефіцієнти регресії.

У разі виявлення зниження ефективності системи технічного обслуговування, здійснюються заходи для її підвищення. Це може включати зміни в робочих процедурах або оновлення режимів роботи обладнання. Паралельно проводиться аналіз (формула 5) даних обладнання для виявлення аномалій і несправностей:

$$Z = \frac{(X - \mu)}{\sigma}, \quad (5)$$

де X – контрольована величина, σ - стандартне відхилення, μ – середнє значення.

У разі їх виявлення формуються рекомендації для їх усунення:

Важливою частиною процесу є звітність та документування усіх виявлених даних та вжитих заходів. Завершується процес актуалізацією параметрів прескриптивної системи технічного обслуговування на основі отриманих результатів (6) за методом машинного навчання для оновлення параметрів на основі набору даних.

$$\theta_{new} = \theta_{old} - \alpha \cdot \nabla \mathrm{J}(\theta), \quad (6)$$

де θ – параметри моделі, α - швидкість навчання, $\mathrm{J}(\theta)$ – функція втрат.

Алгоритм (рис. 2) забезпечує комплексний підхід до моніторингу та управління станом технічного обслуговування вантажного судна, що включає виявлення та реагування на потенційні несправності, а також оптимізацію ефективності системи. Блок-схема (рис. 2) містить в собі методики, які дозволяють об'єктивно оцінювати, аналізувати та оптимізувати процеси в системі технічного обслуговування, що сприяє підвищенню надійності, безпеки та ефективності обслуговування вантажного судна.

3. **Розробка алгоритмів оцінювання способів підвищення ефективності системи ТО з можливістю прогнозування показників ефективності системи ТО та параметрів технічного стану СТЗ і К в умовах експлуатації.** Алгоритм інформаційної системи (рис. 3), важливою задачею якого є оцінювання способів підвищення ефективності системи технічного обслуговування та включає прогнозування

показників ефективності системи та параметрів технічного стану суднових технічних засобів, розпочинається з актуалізації інформації з аналітичним центром ShipDiMRO. Далі відбувається збір і обробка даних, після чого вибирається метод апроксимації для аналізу даних (формула 7):

$$y = a_0 + a_1 x + a_1 x^2 + \cdots + + a_n x^n, \quad (7)$$

Застосування обраного методу апроксимації дозволяє оцінити якість отриманих апроксимацій (формула 8) та визначити, чи знайдена оптимальна модель. Якщо модель вважається оптимальною, обирається результуючий ряд для подальшого аналізу:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} = 1 - \frac{SSE}{SST}, \quad (8)$$

На наступному етапі оцінюється адекватність та оптимальність структури моделі, виділяються статистичні характеристики моделі та розраховується сума квадратів відхилень. Це дозволяє коригувати структуру моделі в разі виявлення значних відхилень:

$$SSE = \sum_i (y_i - \hat{y}_i)^2, \quad (9)$$

Для прогнозування стану системи спочатку встановлюється глибина прогнозування і прогнозний часовий проміжок. Після цього застосовується модель для прогнозування, оцінюється достовірність прогнозу (формули 10 – 13) та формується звіт прогнозу. У разі виявлення недостовірності прогнозу проводиться корекція моделі та повторний аналіз.

За m кроків прогнозування розраховуються такі показники точності прогнозів: середня квадратична похибка (формула 10), корінь із середньоквадратичної похибки (формула 11), середня абсолютна похибка (формула 12), середня абсолютна похибка у відсотках (формула 13):

$$MSE = \frac{\sum_{i=n-m+1}^{n} (y_i - \hat{y}_i)^2}{m}, \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=n-m+1}^{n} (y_i - \hat{y}_i)^2}{m}}, \quad (11)$$

$$MAE = \frac{\sum_{i=n-m+1}^{n} |y_i - \hat{y}_i|^2}{m}, \quad (12)$$

$$MAPE = \sum_{i=n-m+1}^{n} \frac{100 \cdot |y_i - \hat{y}_i|}{m|y_i|}, \quad (13)$$

Якість прогнозу тим вища, чим менше значення величин обчислених за формулами (10 – 13). Даний підхід дає якісні результати, якщо на періоді прогнозу не виникають принципово нові закономірності.

На останньому етапі аналізуються дані звіту прогнозування, ідентифікуються критичні параметри та визначаються часові рамки, протягом яких параметри можуть вийти за встановлені межі. Це дозволяє визначити необхідні заходи реагування та планувати технічне обслуговування.

У разі відсутності критичних відхилень здійснюється моніторинг стану системи.

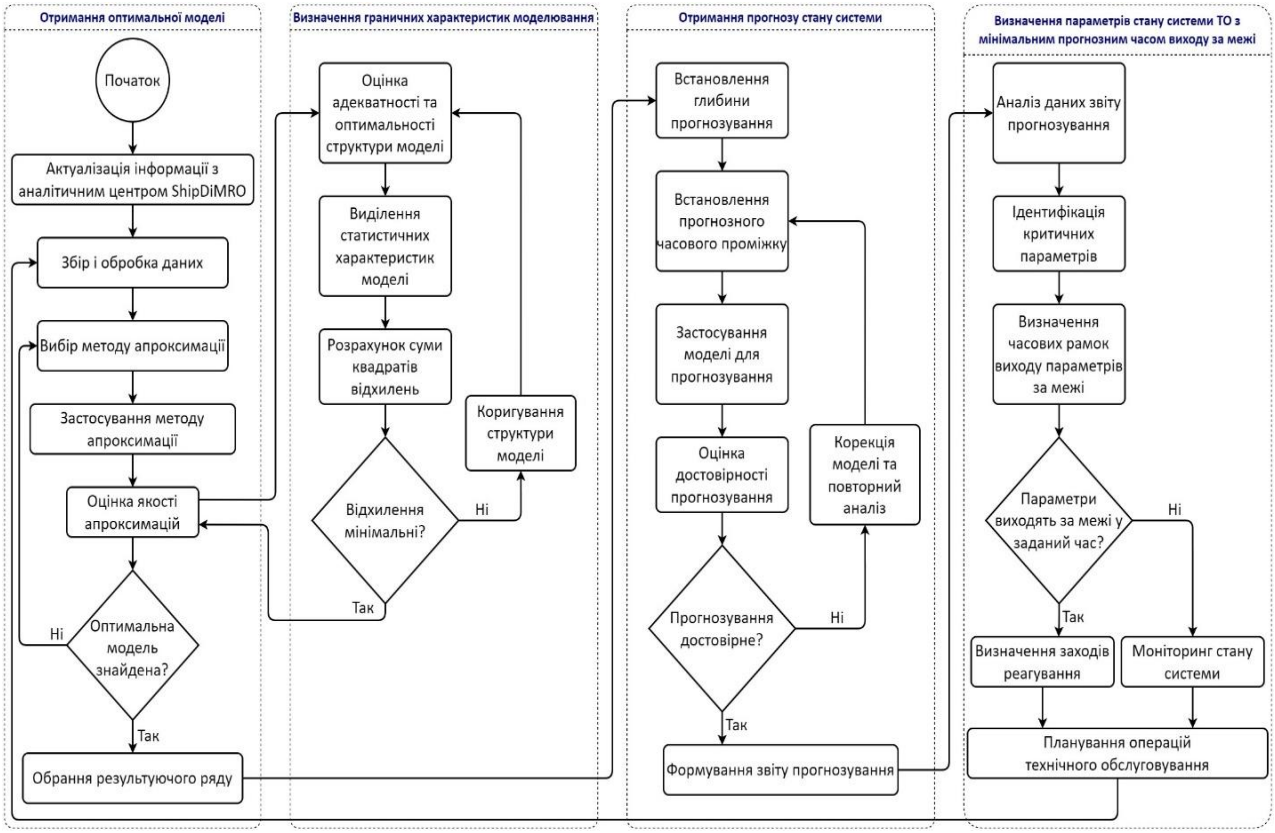


Рис. 3. Блок-схема алгоритму інформаційної системи оцінювання способів підвищення ефективності системи технічного обслуговування з можливістю прогнозування показників ефективності системи технічного обслуговування та параметрів технічного стану суднових технічних засобів і конструкцій в умовах експлуатації

Алгоритм (рис. 3) забезпечує комплексний підхід до оцінювання ефективності системи технічного обслуговування та прогнозування стану технічних засобів, враховуючи поточні тенденції та можливі майбутні зміни.

Висновки

З урахуванням вище зазначених розроблених блок-схем алгоритмів, можна сформулювати такі висновки:

1. Розроблено алгоритм збору даних про параметри системи технічного обслуговування та технічного стану суднових технічних засобів і конструкцій в межах комплексної системи прескриптивного технічного обслуговування.

2. Створено алгоритм моніторингу і визначення статусу несправностей системи технічного обслуговування вантажного судна в межах комплексної системи прескриптивного технічного обслуговування.

3. Представлена блок-схема алгоритму інформаційної системи оцінювання способів підвищення ефективності системи технічного обслуговування з можливістю прогнозування показників ефективності системи технічного обслуговування та параметрів технічного стану суднових технічних засобів і конструкцій в умовах експлуатації.

В цілому, стаття акцентує увагу на необхідності розробки та впровадження комплексних, автоматизованих систем моніторингу для оцінки ефективності технічного обслуговування вантажних суден.

Використання інформаційно-комунікаційних технологій є ключовим для підтримки надійності та ефективності у сфері обслуговування суден.

Стаття підкреслює важливість систематичного підходу, що включає використання оперативних даних та передових аналітичних методів, для покращення процесів обслуговування, забезпечуючи сталу продуктивність та безпеку в морських операціях.

Список літератури

1. Igder, M., Rafiei, M., Boudjadar, J., & Khooban, M. (2021). Reliability and Safety Improvement of Emission-Free Ships: Systemic Reliability-Centered Maintenance. IEEE Transactions on Transportation Electrification, 7, 256-266. https://doi.org/10.1109/TTE.2020.3030082.
2. Bayer, D., Aydin, O., & Çelik, M. (2018). AN ICOR APPROACH TOWARDS SHIP MAINTENANCE SOFTWARE DEVELOPMENT. International Journal of Maritime Engineering. https://doi.org/10.3940/RINA.IJME.2018.A1.444.
3. Anantharaman, M., Khan, F., Garaniya, V., & Lewarn, B. (2014). A Step by Step Approach for Evaluating the Reliability of the Main Engine Lube Oil System for a Ship's Propulsion System. TransNav: International Journal on Marine Navigation and Safety of Sea Transportation, 8, 367-371. https://doi.org/10.12716/1001.08.03.06.
4. Prioletti, M., & Tobin, E. (2008). Automating the Surface Force Tank and Void Assessment Process. Naval Engineers Journal, 120, 53-60. https://doi.org/10.1111/J.1559-3584.2008.00123.X.
5. Sun, C., Wang, H., Liu, C., & Zhao, Y. (2020). Real Time Energy Efficiency Operational Indicator (EEOI): Simulation Research from the Perspective of Life Cycle Assessment. Journal of Physics: Conference Series, 1626. https://doi.org/10.1088/1742-6596/1626/1/012060.
6. Radonjić, A. (2011). Strategy to reduce pollution from Serbian pushboats. International Journal for Traffic and Transport Engineering, 1.
7. Hein, K., Xu, Y., Wilson, G., & Gupta, A. (2020). Condition-based Optimal Maintenance and Energy Management of All-electric Ships. IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, 3767-3772. https://doi.org/10.1109/IECON43393.2020.9254842.
8. Mo, B., Dehli, P., Steinebach, C., Lim, T., & Perera, L. (2017). Automated System for Fleet Benchmarking and Assessment of Technical Condition. . https://doi.org/10.1115/OMAE2017-61219.
9. Anantharaman, M., Islam, R., Khan, F., G., & Lewarn, B. (2019). Data Analysis to Evaluate Reliability of a Main Engine. TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation. https://doi.org/10.12716/1001.13.02.18.
10. Bukša, A., Šegulja, I., & Tomas, V. (2010). Adjustment of Maintenance Approach for Improved Operability and Safety of Ship Navigation. Promet-traffic & Transportation, 22, 95-103. https://doi.org/10.7307/PTT.V22I2.168.

Received (Надійшла) 23.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Features of evaluating the efficiency of cargo vessel maintenance systems

Andrii Golovan

Abstract. The article deals with the important issue of assessing the efficiency of cargo ship maintenance systems. Analysis of recent research and publications in this area indicates the need to assess the efficiency of cargo ship maintenance. The purpose of the article is to develop and form algorithms for an information system for monitoring the efficiency of cargo ship maintenance indicators for use in an information system for assessing ways to improve the efficiency of maintenance. The results of the article include flowcharts of the algorithms for: collecting data on the parameters of the maintenance system and the technical condition of ship's equipment and structures, monitoring and determining the status of faults and evaluating ways to improve the efficiency of the maintenance system with the possibility of forecasting the efficiency of the maintenance system. The conclusions emphasize the importance of analyzing operational data, using information and communication technologies and a systematic approach to evaluating ways to improve efficiency in order to ensure the reliability and efficiency of cargo ship maintenance.

Keywords: maintenance, cargo ships, efficiency, forecasting, analysis.

УДК 629.735.45.066: [004.94:004.93’1] **doi: 10.26906/SUNZ.2024.1.011**

В. В. Коломієць

Державний НДІ випробувань і сертифікації озброєння та військової техніки, Черкаси, Україна

СТРУКТУРА УЗАГАЛЬНЕНОЇ МОДЕЛІ ВІЗУАЛЬНОГО СПОСТЕРЕЖЕННЯ ТА РОЗПІЗНАВАННЯ ОБ’ЄКТІВ, ЯКІ ЗНАХОДЯТЬСЯ В ЗАКАБІННОМУ ПРОСТОРІ ВЕРТОЛЬОТУ

Анотація. Однією з важливих тенденцій розвитку сучасних приладів нічного бачення є покращення їх функціональних можливостей та якості зображення. В статті проаналізовано підґрунтя для розробки математичної моделі візуального спостереження за допомогою приборів нічного бачення, що дозволяють алгоритмізувати методику випробувань окулярів нічного бачення і підвищити ефективність їх використання. Розроблено структури моделі візуального спостереження і моделі об’єкта спостереження в закабінному просторі вертольота, метою яких являється підвищення ефективності наземних випробувань вертольотів та оцінка можливостей льотчика в ергатичній системі “льотчик - вертоліт - оточуюче середовище” здійснення спостереження та розпізнавання об’єктів з використанням окулярів нічного бачення. Розглянуто основні зорові відчуття, які виникають у льотчика, та результати відображення в свідомості спостерігача окремих властивостей об’єкта спостереження при його розпізнаванні.

Ключові слова: авіаційні окуляри нічного бачення, модель візуального розпізнавання, об’єкти спостереження.

Вступ

Постановка проблеми. Необхідну для пілотування вертольоту інформацію (близько 90%), льотчик отримує через зоровий аналізатор [1-5]. Авіаційні окуляри нічного бачення (далі – ОНБ) розроблені для покращення видимості в умовах низької освітленості або повної темряви. Вони використовуються в авіації, де нічний час або умови обмеженої видимості можуть ускладнити польоти. ОНБ засновані на технології інфрачервоного бачення, яка дозволяє отримувати зображення з використанням інфрачервоного випромінювання. Ці окуляри збирають слабке випромінювання, яке є присутнім в темряві, і перетворюють його в зображення, яке може бути сприйняте людським оком.

ОНБ дозволяють пілотам бачити об'єкти, які недоступні для звичайного ока в умовах низької освітленості. Вони допомагають виявляти небезпеку, орієнтуватися в просторі, розрізняти контури та деталі об'єктів. Це може бути особливо важливо під час нічних польотів, при посадці або маневрах у темряві. Однак, при застосуванні ОНБ виникають проблемні питання щодо розпізнавання, відстеження та супроводження цілей та швидкого реагування на них, створення математичної моделі візуального спостереження об’єктів дає змогу підвищити ефективність наземних випробувань вертольотів.

Аналіз останніх досліджень і публікацій. Авіаційні ОНБ Оптико-електронні системи (далі – ОЕС) дозволяють виявляти, розпізнавати, ідентифікувати та визначати координати цілей незалежно від часу доби та застосування противником засобів маскування видимого діапазону оптичного спектра. Поділяють на прилади нічного бачення (далі – ПНБ), тепловізійні прилади, комбіновані (мають декілька каналів на одній платформі, але з роздільними об’єктивами), а також прилади з сумісними каналами зі спостереженням через загальний об’єктив.

У роботах А.В. Луізова [5] розглядається біологічна роль інерції зору, даються відомості про влаштування та роботу ока-приймача, інформації, що отримується людиною за допомогою світла, визначається залежність зорових функцій від світлової обстановки. Розглядаються закони змішування кольорів; колірні простори; колірний тон та частота, розрахунок кольору тощо. Утворився навіть цілий розділ науки, що одержав назву “офтальмоергономіка”, але розрахована вона переважно на біологів і фізіологів.

Треба підкреслити що якість зображення ПНБ визначається класом використаного електронно оптичного перетворювача (далі - ЕОП) та ступенем корекції аберацій оптичних компонентів об’єктива і окуляра. На відміну від поширених ЕОП другого покоління, ЕОП третього покоління принципово відрізняються від своїх попередників високоефективним напівпровідниковим фотокатодом, виготовленим на основі арсеніду галію (AsGa) [3]. Вони мають більшу чутливість та загальний коефіцієнт підсилення яскравості. Через це часто ПНБ, розроблені на їх основі, не потребують додаткового ІЧ-підсвічування. Тому третє покоління ЕОП цінується серед військових та співробітників силових структур [4]. На сьогодні, ПНБ, оснащені такими ЕОП, здатні працювати при низьких рівнях освітленості – до 10^{-5} лк (нічне небо, затягнуте хмарами) [8].

Покращення функціональних можливостей та якості зображення є однією з важливих тенденцій розвитку сучасних приладів нічного бачення. Значний внесок в теорію та практичне оцінювання технічних показників ПНБ належить Маслову С.В. та Есеву А.А. [10].

Глобальна оптимізація як спосіб розв'язання складних задач все частіше використовується у фізиці, техніці, біології, економіці та інших галузях людської діяльності.

В роботах Сокуренко В.М. запропоновано здійснювати автоматизацію розрахунків окулярів для синтезу нових оптичних систем на основі сучасного методу глобальної оптимізації – адаптивного методу диференційної еволюції Коші та синтезовано дві оптичні системи окулярів, які містять дифракційні поверхні та мають практично дифракційно-обмежену якість зображення [6-9, 12-15].

Мета статті: розробка структури моделі візуального спостереження і моделі об’єкта спостереження (далі - ОС) в закабінному просторі вертольота для підвищення ефективності наземних випробувань вертольотів та оцінки можливостей льотчика в ергатичній системі “льотчик - вертоліт - оточуюче середовище” здійснювати спостереження та розпізнавання об’єктів з використанням ОНБ.

Виклад основного матеріалу

Моделі візуального спостереження - це програмні або апаратні системи, які призначені для аналізу та інтерпретації візуальної інформації з використанням обладнання для збору та обробки зображень. Ці моделі можуть виявляти об'єкти, рух та інші важливі деталі на зображеннях або в потоковому відео.

Моделі візуального спостереження мають широкі можливості для покращення безпеки, продуктивності та якості життя у багатьох галузях нашого суспільства. У військовій сфері вони використовуються для розпізнавання, відстеження та супроводження цілей.

Зі зростанням обчислювальної потужності та розвитком штучного інтелекту, моделі візуального спостереження стають все більш потужними та точними.

Це відкриває нові можливості для інновацій та вдосконалення систем, які полегшують ефективність наземних випробувань вертольотів, обладнаних ОНБ та оцінки можливостей льотчика.

На рис. 1 наведена структура узагальненої моделі візуального спостереження:

- модель об’єкта спостереження;
- модель оптичної системи зорового аналізатора (очей);
- модель сітківки зорового аналізатора;
- модель розпізнавання об’єкта спостереження.

Рис. 1. Структура узагальненої моделі візуального спостереження в закабінному просторі вертольота

Кожний ОС (множина матеріальних об'єктів, які перебувають у просторово-часових та інших відношеннях) характеризується променистим потоком (дифузним відображенням світла), розподіленим в просторі та часі (кількість фотонів в одиницю часу):

$$\Phi(П,t) = B_{П}(П,t)\Omega_{П}S_{П}\cos\alpha_{П}, \qquad (1)$$

де $B_{П}(П,t)$ – енергетична яскравість (променистість) ОС; $\Omega_{П}$ – тілесний кут випромінювання поверхні ОС; $S_{П}$ – площа випромінюючої поверхні ОС; $\alpha_{П}$ – кут між напрямом випромінюванням та нормаллю до випромінюючої поверхні.

Причиною випромінювання ОС є наявність відбитого від Місяця світла Сонця і тепловий обмін з навколишнім середовищем. Освітленість від Місяця на поверхні Землі може досягати 0,25 лк, а спектральний діапазон його світла становить 0,4...4 мкм.

Власне випромінювання ОС, пов'язане з тепловим обміном, має максимум випромінювання (λ_{max}, м) у відповідності з законом Віна:

$$\lambda_{\max} = \frac{2898}{T}, \qquad (2)$$

де T – абсолютна температура випромінюючої поверхні (в Кельвінах).

Випромінювання ОС можна описати залежністю:

$$F = \sigma T^4, \qquad (3)$$

де F – потужність на одиницю площі поверхні випромінювання; σ – постійна Стефана-Больцмана.

Розподіл випромінювання ОС по спектру підпорядковується закону Планка ($r_{\lambda,T}$,Вт/м^2):

$$r_{\lambda,T} = c_1\lambda^{-5}\left[e^{c_0/\lambda T} - 1\right]^{-1}, \qquad (4)$$

де c_1 – енергетична яскравість (променистість) ОС; λ – тілесний кут випромінювання поверхні ОС; e – площа випромінюючої поверхні ОС; c_0 – кут між напрямом випромінюванням та нормаллю до

випромінюючої поверхні; T – абсолютна температура випромінюючої поверхні.

Основними характеристиками моделі оптичної системи зорового аналізатора є освітленість на сітківці ока та роздільна здатність ока.

Освітленість на сітківці ока (E_c, лк) розраховується як:

$$E_c = \tau_A \tau_\Gamma S_\Gamma \frac{E_\Gamma}{f_\Gamma^2}; \qquad \alpha = \frac{120^\circ}{D_\Gamma}, \tag{5}$$

де α – роздільна здатність ока; E_Γ – освітленість на вході; S_Γ – площа вхідної зіниці; D_Γ – діаметр вхідної зіниці; f_Γ – фокусна відстані вхідної зіниці; τ_Γ – коефіцієнт пропускання потоку оптикою ока; τ_A – коефіцієнт пропускання потоку середовищем.

У середньостатистичної людини поле зору одного ока складає: по горизонту - 150°, по вертикалі - 125° (дані вірні тільки для ахроматичного зору). Сітківка зорового аналізатора (ока) представляє собою сферичну поверхню, на якій розташовані рецептори: колбочки які відповідають за денний зір а палички відповідають за нічний зір [5].

Рецептори функціонально поєднуються в рецептивні поля які виконують квантування світлового потоку за рахунок статистичної обробки квантів з шумів відбувається виділення корисного сигналу. З концентричних рецептивних полів в нейронний канал видаються для обробки сигнали про освітленість, координати поля, спектри випромінювання (колбочки).

На найбільш високому рівні обробки в сітківці концентричні гангліозні клітини порівнюють освітленість та координати полів, виділяючи контрасти.

Формування зображень і подальше формування образу ОС пов'язаний з подоланням невизначеностей у зв'язку з можливістю використання метричних співвідношень при проективному відображенні, окремим випадком якого є центральна проекція і випадковістю процесу відображення [16].

Складність розпізнавання просторових об'єктів при спостереженні їх довільними ракурсами полягає у виявленні таких ознак, які інваріантні не тільки при ізоморфних перетвореннях (зміна масштабу, положення об’єкту тощо), але і при проектному відображенні.

Вважаємо, що деяке рецептивне поле – детектор реєструє проходження через нього межі зображення і визначає кут нахилу цієї межі до деякого фіксованого напрямку, допускаючи, що розміри поля настільки малі, що кордон є прямолінійним [17].

Матриця таких детекторів дозволяє виділити зображення у вигляді відрізків, сполучених у точках зламу-вершинах.

Вважаючи межі ребрами, приходимо до розгляду плаского неорієнтованого графа.

Втім кількість графів-зображень, які відповідають образу ОС, нескінченно [18]. Користуючись поняттям ізоморфізму графів, вдається звести нескінченну множину зображень до кінцевої множини ізоморфних графів. Характеристикою графів є кількість вершин у співвідношенні до опуклих та увігнутих вершин, послідовність розташування цих вершин по контуру тощо.

Вважаємо, що механізм розпізнавання зображень по формі в зоровому аналізаторі схожий до наведеного вище [16-18].

Для підтвердження цієї гіпотези розглянемо ймовірність розпізнавання ОС за формою [18,20]:

$$p = \exp\left[-F\left(\frac{A}{L}\right)^2\right], \tag{6}$$

де F – коефіцієнт форми; A – роздільна здатність системи спостереження на місцевості; L – розмір об'єкту спостереження.

Значення коефіцієнтів форми, отримані експериментально та теоретично, представлені в табл. 1 [18].

Таблиця 1 – **+Значення коефіцієнтів форми**

F	**Форми об’єктів спостереження**					
	□	○	⌐	▭	△	⬡
$F_{експ}$	1,72	0,97	1,58	2,78	-	-
F_{T1}	1,55	1,42	1,60	2,61	1,74	1,47
F_{T2}	1,65	1	1,65	2,73	1,93	1,39

Значення F_{T1} розраховують за формулами [18]:

$$F_{T1} = \sqrt{\frac{G\overline{R}}{S}}; \qquad \overline{R} = \frac{R_П + R_0}{2}, \tag{7}$$

де G – периметр по контуру зображення; $R_П$ – радіус кола, вписаною в зображення; R_O – радіус кола, описаною навколо зображення; S – площа зображення.

В основу обчислення коефіцієнта та форми F_{T1} та F_{T2} положенні об'єктивно існуючі співвідношення між периметром і площею геометричних фігур, їх залежність від масштабу, представлення про механізм виділення цих метричних ознак у вигляді концентричних полів.

Значення F_{T2} обчислювалося за формулою:

$$F_{T2} = e^{\frac{2+b_v}{b}}, \tag{8}$$

де b – кількість вершин в графі – зображення; b_v – кількість увігнутих вершин.

З табл. 1 видно задовільний збіг значень F_{T2} до $F_{експ}$. Крім того, при обчислені не використовуються метричні співвідношення. Додатково можна відмітити простоту виразу для розрахунку значення F_{T2}, що також свідчить на користь запропонованої гіпотези (методики).

Розглянемо основні зорові відчуття, які виникають у льотчика, та результати відображення в свідомості спостерігача окремих властивостей ОС при його розпізнаванні, які представлені на рис. 2.

Рис. 2. Структура властивостей об'єкта спостереження

Світловідчуття характеризується мінімальною освітленістю на зіниці, що викликає почуття світла [5]. Порогова освітленість денного зору складає 7×10^{-8}лк при темному фоні нічний зір залежить від площі рецептивного поля: при

$$S_{max}=15,\ E_{пор}=1{,}5\times10^{-9}\ \text{лк}$$

при темному фоні.

Прямий контраст (відчуття контрасту за яскравістю)обчислюється за формулою:

$$K_{ПР}=\frac{B_Ф-B_O}{B_O}, \qquad (9)$$

де B_O, $B_Ф$ – яскравість об'єкта і фону.

Відчуття руху виникає при зміщенні зображення з рецептивного поля.

Порогова швидкість переміщення точкового зображення $\omega_{пор}$ визначається часом формування світловідчуття (t_c) та такою роздільною здатністю сітківки α:

$$\omega_{пор}=\frac{\alpha}{t_c}. \qquad (10)$$

Час, необхідний для виникнення зорового відчуття, залежить від яскравості об'єкта і довжини хвилі випромінювання й змінюється в межах 0,02 ... 0,1 с.

В одиночному зоровому акті для виключення флуктуації потоку необхідно прийняти певну кількість фотонів.

Підвищення гостроти зору означає зменшення площі приймача, що компенсується збільшенням часу підсумовування фотонів.

При

$$t_c = 0{,}02 \text{ с та } \alpha = 1'$$

отримуємо порогове значення кутової швидкості точкового джерела порядку 1°/с [5].

Висновки

Модель візуального розпізнавання об'єктів, яка знаходиться в закабінному просторі вертольота, може бути використана для різних цілей і завдань, в залежності від конкретних потреб операторів вертольота що дозволяє приймати швидкі та відповідні рішення для забезпечення безпеки польотів, навігації та виконання різних завдань.

Використання моделі візуального розпізнавання об'єктів при формуванні алгоритмів забезпечення наземних випробувань бойових вертольотів, обладнаних авіаційними окулярами нічного бачення, дозволить підвищити якість і достовірність випробувань, а також буде сприяти скороченню термінів, обсягів та витрат на їх проведення.

СПИСОК ЛІТЕРАТУРИ

1. Коломієць В.В. Аналіз особливостей використання окулярів нічного бачення пілотом вертольоту та їх впливу на безпеку польотів /В.В. Коломієць // Системи управління, навігації та зв'язку. – 2023. – №1. – С. 36-39. DOI: https://doi.org/10.26906/SUNZ.2023.1.036
2. Неня О.В. Сучасні теплові зори для спеціального та повсякденного застосування / Сучасна спеціальна техніка. 2016. № 4(47)- С. 212-259. Режим доступу https://elar.naiau.kiev.ua/server/api/core/bitstreams/11475cb8-6431-41b6-8dea-a13063954eae/content

3. Розрахунок і конструювання оптико-електронних приладів : навч. посібник / А. С. Литвиненко, Г. О. Петченко, О. М. Ляшенко, О. М. Діденко ; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. – Харків : ХНУМГ ім. О. М. Бекетова, 2021. – 139 с. ISBN 978-966-695-558-9
4. Guterman P. S. Assessing night vision goggle performance in security applications [Electronic resource] / P. S. Guterman // Academia. – Режим доступу https://yorku.academia.edu/PearlGuterman (дата звернення: 21.08.2023). – Назва з екрана.
5. Луизов А. В. Глаз и свет : / А. В. Луизов. - Л. : Энергоатомиздат, 1983. – С. 139, [5]. :табл. - Библиогр.: С. 144.
6. Сокуренко В. М., Макаренко Я. І. Розробка оптичних систем методами глобальної оптимізації // Вісник НТУУ "КПІ". Серія приладобудування. – 2015. – № 50(2). – С. 51-60.
7. Режим доступу https://ela.kpi.ua/handle/123456789/16144
8. Сокуренко В. М., Вакуленко М. М. Автоматизований розрахунок окулярів з дифракційними оптичними елементами // Вісник Хмельницького національного університету: Технічні науки. – Хмельницький. – 2018. – №1 (257). – С. 107-112. Режим доступу http://journals.khnu.km.ua/vestnik/pdf/tech/pdfbase/2018/2018_1/jrn/pdf/21.pdf
9. Сокуренко В. М., Бондарчук Д. П. Автоматизований параметричний синтез фотооб'єктива зі зменшеною дисторсією // Вісник НТУУ "КПІ". Серія приладобудування. – 2018. – № 56(2). – С. 18-24. Режим доступу https://ela.kpi.ua/handle/123456789/24142
10. Проскурін В.С., Сокуренко В.М. Автоматизований розрахунок оптичної системи приладу нічного бачення зі збільшеним кутом огляду // XIV Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених «Погляд у майбутнє приладобудування», 18-19 травня 2021 року, : Оптичні та оптико-електронні прилади і системи.– Київ: КПІ ім. Ігоря Сікорського, 2021 – С. !22-125. Режим доступу https://ela.kpi.ua/bitstream/123456789/46727/1/Page122-125.pdf
11. Clark G. Helicopter Handling Qualities in the Degraded Visual Environment / G. Clark. – DVEY London : University of Liverpool, 2003. Режим доступу https://livrepository.liverpool.ac.uk/id/eprint/3174810
12. Микитенко В.І. Порівняння якості роботи фільтрів для зменшення шумів зображень тепловізійного каналу оглядових оптико-електронних пристроїв / Балахонова Н. О., Микитенко В. І., Пашков Р. А. // Вісник КПІ. Серія приладобудування. - 2019. - Вип. 57(1). - С.26 – 35. Режим доступу http://nbuv.gov.ua/UJRN/VKPI_prylad_2019_57%281%29__6
13. Сокуренко В.М. Числове дослідження стохастичних методів безперервної глобальної оптимізації / В. М. Сокуренко, В. С. Неділюк // Наукові вісті Національного технічного університету України "Київський політехнічний інститут". - 2012. - № 1. - С. 81-88. Режим доступу http://nbuv.gov.ua/UJRN/NVKPI_2012_1_12
14. Сокуренко В. М., Макаренко Я. І. Розробка оптичних систем методами глобальної оптимізації // Вісник НТУУ "КПІ". Серія приладобудування. – 2015. – № 50(2). – С. 51-60. Режим доступу https://ela.kpi.ua/handle/123456789/16144
15. Сокуренко В. М., Вакуленко М. М. Автоматизований розрахунок окулярів з дифракційними оптичними елементами // Вісник Хмельницького національного університету: Технічні науки. – Хмельницький. – 2018. – №1 (257). – С. 107-112. Режим доступу http://journals.khnu.km.ua/vestnik/pdf/tech/pdfbase/2018/2018_1/jrn/pdf/21.pdf
16. Сокуренко В. М., Тростянська О. В. Синтез оптичної системи окуляра для мікродисплея з високою роздільною здатністю // Вісник Хмельницького національного університету: Технічні науки. – Хмельницький, 2019. – №6(279). – с. 206-210. DOI 10.31891/2307-5732-2019-279-6-206-210
17. Рішення Dahua Full-color AI для відеокамер зі штучним інтелектом: робить ніч такою ж барвистою, як і день / https://www.bezpeka-shop.com/ua/blog/obzor/reshenie-dahua-full-color-ai-dlya-videokamer-s-iskusstvennym-intellektom
18. Назаренко Л. А. Основи радіометрії та фотометрії: монографія / Л. А. Назаренко, В. М. Сорокін; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. – Харків : ХНУМГ ім. О. М. Бекетова, 2014. – 352 с. Режим доступу https://core.ac.uk/download/pdf/187144168.pdf
19. Tatyanko D. N. Quantum efficiency improvement of optical radiation trap-detectors / D. N. Tatyanko, P. I. Neyezhmakov, Ye. P. Timofeev, A. S. Litvinenko, K. I. Suvorova, O. M. Didenko // Semiconductor physics, quantum electronics & optoelectronics. - 2019. - Vol. 22, № 1. - С. 104-110. DOI: https://doi.org/10.15407/spqeo22.01.104
20. Микитенко В.І. Ефективність інфрачервоних оптико-електронних систем спостереження: монографія / В.Г. Колобродов, В.І. Микитенко, Є.Г. Балінський // Київ: «Вік прінт», 2017. - 202 с. Режим доступу https://ela.kpi.ua/handle/123456789/26726
21. Микитенко В.І. Обґрунтування параметрів оптико-електронної системи спостереження для мікросупутників / Тимчик Г.С., Колобродов В.Г., Микитенко В.І. // Перша Всеукраїнська конференція «Аерокосмічні спостереження в інтересах сталого розвитку та безпеки» (3-5 червня 2008 р.). Збірник тез доповідей / КП СПБ «Арсенал». - Київ, 2008. - С. 47 - 48.

Received (Надійшла) 30.11.2023
Accepted for publication (Прийнята до друку) 24.01.2024

Structure of the generalized model visual observation and recognition of objects, which are located in the cabin space of the helicopter

Volodymyr Kolomiiets

Abstract. One of the important trends in the development of modern night vision devices is the improvement of their functionality and image quality. The article analyzes the basis for the development of a mathematical model of visual observation with the help of night vision devices, which allows to algorithmize the methodology of testing night vision glasses and increase the efficiency of their use. The structures of the model of visual observation and the model of the object of observation in the cockpit space of the helicopter were developed, the purpose of which is to increase the efficiency of ground tests of helicopters and to assess the capabilities of the pilot in the energetic system "pilot - helicopter - environment" to observe and recognize objects using night vision goggles vision. The basic visual sensations experienced by the pilot and the results of the reflection in the mind of the observer of certain properties of the object of observation during its recognition are considered.

Keywords: aviation night vision goggles, visual recognition model, surveillance objects.

УДК 623.746.2 **doi: 10.26906/SUNZ.2024.1.016**

Р. В. Куліш

Льотна академія Національного авіаційного університету, Кропивницький, Україна

МОДЕЛЬ УРАХУВАННЯ БІЧНОГО ВІТРУ ПРИ ПЛАНУВАННІ МАРШРУТІВ ПОЛЬОТУ БЕЗПІЛОТНОГО ЛІТАЛЬНОГО АПАРАТУ

Анотація. В статті ґрунтуючись на сучасних методах оптимізації траєкторія польоту безпілотного літального апарату (БпЛА), що враховують існуючи обмеження, розроблена модель врахування бічного вітру при плануванні маршрутів польоту БпЛА, яка ґрунтується на побудові оптимального фільтра Калмана. Дана модель дозволяє корегувати маршрутні точки БпЛА при виконанні завдань моніторингу елементів критичної інфраструктури, що зменшує маршрут польоту БпЛА при виконанні завдання моніторингу з урахуванням складних умов польоту та зменшує ступінь завантаження бортового обчислювального пристрою БпЛА.

Ключові слова: безпілотний літальний апарат, маршрутизація, моніторинг об’єкти критичної інфраструктури, планування маршрутів.

Вступ

Постановка проблеми. Розв’язана росією повномасштабна війна проти України вимагає концентрації зусиль всього суспільства та наукової думки щодо протидії агресору. Об’єкти критичної інфраструктури стали основними цілями для нанесення ударів державою-терористом особливо в зимовий період [1, 2]. Генеруючі потужності, електропідстанції, високовольтні лінії електропередач, об’єкти комунальної інфраструктури міст вимагають постійного моніторингу та оперативного реагування як в умовах війни, так і в умовах мирного часу.

Для отримання оперативної та достовірної інформації щодо елементів критичної інфраструктури (ЕКІ) можна активно застосовувати безпілотні літальні апарати [3, 4].

Вони здатні здійснювати:

- моніторинг стану транспортних магістралей та мостів, нафто- й газопроводів, ліній електропередач, елементів\ електростанцій та інших ЕКІ;
- проводити інженерну розвідку районів повеней, землетрусів та інших стихійних лих на об’єктах критичної інфраструктури, визначення точних координат постраждалих ЕКІ;
- доставку малогабаритних спеціальних вантажів та засобів, необхідних для проведення аварійно-рятувальних робіт й життєзабезпечення персоналу об’єктів критичної інфраструктури при терористичних атаках;
- виявлення ступеня хімічного (на хімічно небезпечних об'єктах) або радіоактивного (на радіаційних небезпечних об’єктах) зараження місцевості зі встановленням точних даних про концентрацію шкідливих речовин й рівень небезпечного випромінювання для визначення можливості спрямування рятувальників, вибору часу й режиму їх роботи, а також необхідних засобів індивідуального захисту;
- патрулювання територію розташування об’єкта критичної інфраструктури з метою запобігання несанкціонованому доступу на цю територію;
- моніторингу ЕКІ, отримання оперативної інформації щодо їх стану.

Рішення завдання ефективного планування для моніторингу ЕКІ має свої особливості, що зумовлені великою площею пошуку, незначними (або навпаки, значними) геометричними розмірами об’єктів моніторингу, необхідністю розпізнавання критичних змін, пошуку динамічних об’єктів (як, наприклад, порушників периметру) об’єкту, несприятливими факторами зовнішнього середовища для виконання польотів, необхідністю вибору доцільного корисного навантаження для БпЛА. Розроблення методів моніторингу ЕКІ за допомогою БпЛА вимагає розроблення моделей щодо врахування впливу негативних умов польоту. Одним з факторів, що найбільше впливають на політ БпЛА, є бічний вітер, який не вимірюється більшістю сучасних бортових датчиків недорогих БпЛА. Таким чином, для підвищення точності маршруту необхідно оцінити кут вітрового зносу за допомогою ідентифікатора.

Аналіз останніх досліджень і публікацій. В роботах по плануванню застосування БпЛА [6, 7] визначаються основні несприятливі фактори планування польоту: вітер, опади, хмарність, зледеніння.

В літературі по метеорології вказано [8, 9], що погода постійно змінюється, причому ці зміни можуть бути помітні на досить короткому проміжку часу. Зміни погоди бувають періодичними та неперіодичними. Неперіодичні зміни пов’язані з рухом повітряних мас. Рух повітряних мас викликається нерівномірним прогрівом земної поверхні внаслідок різного кута падіння сонячних променів. Біля екватора наземний повітряний шар прогрівається значно краще, ніж на полюсах, що викликає рух повітряних потоків по всій земній кулі. При перенесенні повітряних мас із одних районів Землі до інших переносяться і характеристики погоди. Тому для фізико-географічних умов конкретної країни можна визначити, який вітер приносить яку погоду. Крім цього, для кожного регіону є типова погода, властива лише цьому регіону. Наприклад, для Києва та пригородів типовий й найчастіший напрямок вітру є західний. Для Луганська типовий вітер – східний. Отже східний вітер у районі Києва свідчить про швидку зміну погоди, а в районі Луганську – навпаки.

Зі збільшенням висоти вітер посилюється через зменшення сили тертя поверхню Землі. Земна пове-

рхня неоднорідна за своїм кольором, рельєфом і вологістю. Тому під впливом сонячних променів вона прогрівається нерівномірно. Темніші і сухі поверхні прогріваються значно швидше, ніж світлі та вологі, віддаючи значно більше тепла повітрю. Також на прогрів значно впливають схили та височини, наприклад південний схил тепліший за північний, а східний схил прогрівається раніше ніж західний. Усі ці чинники створюють передумови виникнення термічної активності повітря. Тепліше, прогріте повітря, спрямовується вгору (виникають висхідні термічні потоки, або терміки), а на його місце підтягується повітря з холодніших зон (виникає термічний, або місцевий вітер). Так виникають посилення чи навпаки стихання вітру у наземному шарі за умови наявності фонового вітру. За відсутності фонового вітру місцевий вітер дме різноспрямовано й регулярно (у середині дня повного штилю немає практично ніколи).

Дослідники погоджуються, що для підвищення ймовірність успіху місії БпЛА, траєкторія польоту повинна бути ретельно розрахована з урахуванням методів оптимізації при існуючих обмеженнях. Незважаючи на те, що алгоритми на основі графів, такі як діаграма Вороного, search A*, D* lite та інші класичні методи, такі як швидке дослідження випадкових дерев (RRT), штучні потенційні поля (APF) та ймовірнісні дорожні карти (PRM), зазвичай використовуються для розрахунку траєкторії БпЛА, вони вимагають створення карт вартості складних полів і, зазвичай, їх недоліком є збіжність до локальних оптимальних рішень [10, 11].

Останніми роками метаевристичні алгоритми, які є розгалуженням методів штучного інтелекту, почали використовуватися як планувальники маршруту груп або окремих БпЛА через їх переваги щодо простоти реалізації, обчислювальної складності та конфігурованих або настроюваних структур [12, 13, 14]. В той же час для звичайних БпЛА відсутні ефективні моделі врахування такого явища, як боковий вітер, що вимагає розроблення відповідної моделі.

Мета статті: розроблення моделі врахування бічного вітру при плануванні маршрутів польоту безпілотного літального апарату.

Виклад основного матеріалу

На підставі наявної вимірювальної інформації проводиться непряме оцінювання координат БпЛА з використанням алгоритмів оптимальної фільтрації на основі фільтра Калмана (ФК) [15, 16].

Алгоритм реалізації фільтра Калмана дозволяє в реальному часі побудувати оптимальну оцінку стану системи, ґрунтуючись на вимірах, які містять похибки; при цьому вектор вимірювань розглядається як багатовимірний вихідний сигнал системи, обтяженого шумом, а вектор стану – невідомий багатовимірний сигнал, що підлягає визначенню. Умовою оптимальності побудованої оцінки стану є мінімум середньої квадратичної помилки.

Проведемо оцінювання вектора стану безпілотного літального апарату. Побудова ідентифікаторів об'єкта управління з використанням оптимального фільтра Калмана (ОФК) вимагає виконання таких умов:

1) вектор стану $x(t)$ повинен задовольняти векторному диференціальному рівнянню:

$$\frac{dx}{dt} = F(t) \cdot x(t) + H(t) \cdot \xi(t), \quad (1)$$

де $x(t)$ – вектор-функція, що містить компонент; $F(t)$ – квадратична матриця розміром $n \times n$, що залежить у загальному випадку від часу; $\xi(t)$– вектор білих шумів, що складаються з l компонент; $H(t)$ - матриця розміром $n \times l$.

2) кореляційна матриця процесу $\xi(t)$ є такою:

$$Q(t,u) = Q(t) \cdot \delta(t-u), \quad (2)$$

де $Q(t)$ – квадратна матриця розміром $l \times l$.

3) сукупність процесів на m виходах фільтра утворює векторний спостережуваний процес:

$$y(t) = C(t) \cdot x(t) + n(t), \quad (3)$$

де $y(t)$ – вектор-функція, що складається з елементів; $C(t)$ –прямокутна матриця розміром $m \times n$; $n(t) - m$ –мірний вектор білих шумів (шумів вимірювання) з кореляційною матрицею:

$$R(t,u) = R(t) \cdot \delta(t-u), \quad (4)$$

де $R(t)$ – позитивно-визначена квадратична матриця розміром $m \times m$.

4) структура ідентифікатора описується векторним диференціальним рівнянням:

$$\frac{d\hat{x}}{dt} = F(t) \cdot \hat{x}(t) K(t,t) \times [y(t) - C \cdot \hat{x}(t)], \quad (5)$$

де $\hat{x}(t)$ – векторний процес на виході оптимального фільтра, який відтворює з деякою помилкою процес $x(t)$:

$$K(t,t) = P(t) - C^T(t) - R^{-1}(t), \quad (6)$$

$P(t)$ – матриця дисперсії помилок фільтрації, зміна якої описується таким рівнянням:

$$\begin{aligned} \frac{dP}{dt} = {} & F(t) \cdot P(t) + P(t) \cdot F^T(t) - \\ & - P(t) \cdot C^T(t) \cdot R^{-1}(t) \cdot C(t) \cdot P(t) + \\ & + H(t) \cdot Q(t) \cdot H^T(t). \end{aligned} \quad (7)$$

Для запобігання старінню коефіцієнтів фільтра та покращення оцінки кута ковзання через деякий час після початку роботи фільтра було зупинене інтегрування коваріаційної матриці P, тобто коефіцієнти фільтра мають бути взяті як константи на деякий час, який залежить від типу БпЛА та сили бічного вітру. Час підбирається експериментальним шляхом і для БпЛА літакового типу складає близько 2 с. Запишемо диференціальні рівняння бічного руху БпЛА при дії бічного вітру.

$$\begin{cases} \dot{\beta} = \bar{Z}^{\beta} + \bar{Z}_{\beta}^{\omega_X} \omega_X + \bar{Z}_{\beta}^{\omega_Y} \omega_Y + \bar{Z}_{\beta}^{\gamma}, \\ \dot{\omega}_X = \bar{M}_x^{\beta} \beta + \bar{M}_x^{\omega_X} \omega_X + \bar{M}_x^{\omega_Y} \omega_Y + \bar{M}_x^{\delta_{\text{Е}}} \delta_{\text{Е}}, \\ \dot{\omega}_Y = \bar{M}_y^{\beta} \beta + \bar{M}_y^{\omega_X} \omega_X + \bar{M}_y^{\omega_Y} \omega_Y + \bar{M}_y^{\delta_{\text{Н}}} \delta_{\text{Н}}, \\ \dot{\gamma} = \omega_X, \\ \dot{\psi} = \hat{\omega}_Y, \\ \dot{\phi} = \bar{Z}_{\beta}^{\psi} \beta + \bar{Z}_{\omega_X}^{\psi} \omega_x^{\omega_X}, \end{cases} \quad (8)$$

де β – кут ковзання БпЛА; ω_X – кутова швидкість обертання БпЛА навколо осі OX; ω_Y – кутова швидкість обертання БпЛА навколо осі OY; γ – кут крену БпЛА; ψ – кут курсу БпЛА; ϕ– кут шляху БпЛА; δ_{E} – кут відхилення елеронів; δ_{H} – кут відхилення рулів напрямку; $\bar{Z}^{\beta}, \bar{Z}_{\beta}^{\omega_X}, \bar{Z}_{\beta}^{\omega_Y}, \bar{Z}_{\beta}^{\gamma}, \bar{M}_x^{\beta}, \bar{M}_x^{\omega_X}, \bar{M}_x^{\omega_Y}, \bar{M}_x^{\delta_{\text{E}}}, \bar{M}_y^{\beta}, \bar{M}_y^{\omega_X}, \bar{M}_y^{\omega_Y}, \bar{M}_y^{\delta_{\text{H}}}$ – коефіцієнти математичної моделі (ММ) бічного руху БпЛА.

Запишемо систему рівнянь (8) у вигляді:

$$\dot{x} = \text{Ax} + \text{B}u, \tag{9}$$

де

$$x = \begin{pmatrix} \beta \\ \omega_X \\ \omega_Y \\ \gamma \\ \psi \\ \phi \end{pmatrix},$$

$$A = \begin{pmatrix} \bar{Z}^{\beta} & \bar{Z}_{\beta}^{\omega_X} & \bar{Z}_{\beta}^{\omega_Y} & \bar{Z}_{\beta}^{\gamma} & 0 & 0 \\ \bar{M}_x^{\beta} & \bar{M}_x^{\omega_X} & \bar{M}_x^{\omega_Y} & 0 & 0 & 0 \\ \bar{M}_y^{\beta} & \bar{M}_y^{\omega_X} & \bar{M}_y^{\omega_Y} & \bar{M}_y^{\delta_{\text{H}}} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \bar{Z}_{\beta}^{\psi} & \bar{Z}_{\omega_X}^{\psi} & 0 & 0 & 0 \end{pmatrix}, \tag{10}$$

$$u = \begin{pmatrix} \delta_{\text{H}} \\ \delta_{\text{E}} \end{pmatrix}, B = \begin{pmatrix} 0 & 0 \\ 0 & \bar{M}_x^{\delta_{\text{E}}} \\ \bar{M}_y^{\delta_{\text{H}}} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}. \tag{11}$$

На рис. 1 наведено перелік вхідних та вихідних координат ідентифікатора, побудованого на базі оптимального ФК.

Рис. 1. Ідентифікатор кута ковзання та кута вітрового зносу: γ,ψ – кути крену та курсовертикалі; ω_X, ω_Y – кутові швидкості навколо пов'язаних осей з датчиків кутових швидкостей; ϕ – кут шляху; $\hat{\beta}$ – оцінка кута ковзання; $\widehat{\beta}_W$ – оцінка кута вітрового зносу

Наведемо рівняння ідентифікатора зі сталими коефіцієнтами для поздовжнього каналу:

$$\hat{\beta} = \bar{Z}^{\beta}\hat{\beta} + \bar{Z}_{\beta}^{\omega_X}\hat{\omega}_X + \bar{Z}_{\beta}^{\omega_Y}\hat{\omega}_Y + \bar{Z}_{\beta}^{\gamma}\hat{y} + K_{\omega_X}^{\beta}(Z_{\omega_X} - \hat{\omega}_X) + K_{\omega_y}^{\beta}\left(Z_{\omega_y} - \hat{\omega}_y\right) + K_{\gamma}^{\beta}(Z_{\gamma} - \hat{\gamma}) + K_{\psi}^{\beta}(Z_{\psi} - \hat{\psi}),$$

$$\hat{\omega}_x = \bar{M}_x^{\beta}\hat{\beta} + \bar{M}_x^{\omega_X}\hat{\omega}_X + \bar{M}_x^{\omega_Y}\hat{\omega}_Y + K_{\omega_X}^{\omega_X}(Z_{\omega_X} - \hat{\omega}_x) + K_{\omega_y}^{\omega_X}(Z_{\omega_X} - \hat{\omega}_y) + $$
$$+K_{\gamma}^{\omega_x}(Z_{\gamma} - \hat{\gamma}) + K_{\psi}^{\omega_x}(Z_{\psi} - \hat{\psi}) + \bar{M}_x^{\delta_{\text{E}}}\delta_{\text{E}},$$

$$\hat{\omega}_y = \bar{M}_y^{\beta}\hat{\beta} + \bar{M}_y^{\omega_X}\hat{\omega}_X + \bar{M}_y^{\omega_Y}\hat{\omega}_Y + K_{\omega_x}^{\omega_y}(Z_{\omega_x} - \hat{\omega}_x) + K_{\omega_y}^{\omega_y}\left(Z_{\omega_y} - \hat{\omega}_y\right) + \tag{12}$$
$$+K_{\gamma}^{\omega_y}(Z_{\gamma} - \hat{\gamma}) + K_{\psi}^{\omega_y}(Z_{\psi} - \hat{\psi}) + \bar{M}_y^{\delta_{\text{H}}}\delta_{\text{H}},$$

$$\hat{\gamma} = \hat{\omega}_X + K_{\omega_X}^{y}(Z_{\omega_X} - \hat{\omega}_x) + K_{\omega_y}^{y}\left(Z_{\omega_y} - \hat{\omega}_y\right) + K_{\gamma}^{\psi}(Z_{\gamma} - \hat{\gamma}) + K_{\psi}^{\psi}(Z_{\psi} - \hat{\psi}),$$

$$\hat{\psi} = \hat{\omega}_Y + K_{\omega_x}^{\psi}(Z_{\omega_x} - \hat{\omega}_x) + K_{\omega_y}^{\psi}\left(Z_{\omega_y} - \hat{\omega}_y\right) + K_{y}^{\psi}(Z_{y} - \hat{\gamma}) + K_{\psi}^{\psi}(Z_{\psi} - \hat{\psi}),$$

$$\hat{\phi} = \bar{Z}_{\beta}^{\psi}\beta + \bar{Z}_{\omega_X}^{\psi}\omega_x^{\omega_X} + \bar{Z}_{\gamma}^{\psi}\gamma + K_{\omega_X}^{\psi}(Z_{\omega_X} - \hat{\omega}_x) + K_{\omega_y}^{\psi}\left(Z_{\omega_y} - \hat{\omega}_y\right) + K_{y}^{\psi}(Z_{y} - \hat{\gamma}) + K_{\psi}^{\psi}(Z_{\psi} - \hat{\psi}).$$

де Z_{ω_X}– виміряна кутова швидкість обертання БпЛА навколо осі OX; Z_{ω_y}– виміряна кутова швидкість обертання БпЛА навколо осі OY; Z_{γ}– виміряний кут крену БпЛА; Z_{ψ}- виміряний кут курсу БпЛА; Z_{ψ}– виміряний шляховий кут БпЛА; $\hat{\beta}$, $\hat{\omega}_X$, $\hat{\omega}_Y$, $\hat{y}$,$\hat{\psi}$, $\hat{\psi}$ – оцінки відповідних параметрів; $\bar{Z}^{\beta}, \bar{Z}_{\beta}^{\omega_X}, \bar{Z}_{\beta}^{\omega_Y}, \bar{Z}_{\beta}^{\gamma}, \bar{M}_x^{\beta}, \bar{M}_x^{\omega_X}, \bar{M}_x^{\omega_Y}, \bar{M}_y^{\beta}, \bar{M}_y^{\omega_X}, \bar{M}_y^{\omega_Y}, \bar{M}_x^{\delta_{\text{E}}}, \bar{M}_y^{\delta_{\text{H}}}$ – коефіцієнти математичної моделі руху БпЛА.

Варто звернути увагу на те, що сам кут вітрового зносу обчислюється алгебраїчно, виходячи з формули:

$$\hat{\beta}_W = \phi - \psi + \hat{\beta}. \tag{13}$$

Моделювання системи управління та ідентифікатора проводилося у середовищі імітаційного моделювання Simulink.

Вітер моделювався у вигляді поривів, тобто зміна кута вітрового зносу була стрибком або дуже “крутою” лінійною залежністю.

Задамось такими значеннями похибками датчиків та середньоквадратичним значенням швидкості поривів вітру:

- похибка датчика кутових швидкостей при вимірі кутової швидкості крену $\sigma_{\omega x}$ = 0,0029 рад/с;
- похибка датчика кутових швидкостей при вимірі кутової швидкості ризику $\sigma_{\omega y}$ = 0,0029 рад/с;
- похибка курсовертикалі при вимірі кута крену $\sigma_{\gamma} = 0{,}0172$ рад ;
- похибка курсовертикалі під час вимірювання кута курсу $\sigma_{\psi} = 0{,}05$ рад;
- похибка супутникової навігаційної системи під час вимірювання кута шляху σ_{ψ}= 0,03 рад;
- середньоквадратичне значення швидкості поривів вітру $\sigma_{\beta_{\text{B}}}^2 = 1(\text{м/хв})^2$.

Тоді матриці шумів вимірювання (R) та вітрового збурення (Q) набудуть вигляду:

$$(Q) = \sigma_{\beta_{\text{B}}}^2,$$

$$R = \begin{pmatrix} \sigma_{\omega_X}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{\omega_Y}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\omega_\gamma}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_\psi^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\psi^2 \end{pmatrix}.$$

Початкова коваріаційна матриця була обрана експериментальним шляхом:

$$P_0 = \begin{pmatrix} 0.02 & 0 & 0 & 0 & 0 \\ 0 & 0.05 & 0 & 0 & 0 \\ 0 & 0 & 0.05 & 0 & 0 \\ 0 & 0 & 0 & 0.02 & 0 \\ 0 & 0 & 0 & 0 & 0.02 \end{pmatrix}.$$

Матриця спостереження БЛА:

$$H = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Матриці F та B аналогічні матрицям з математичної моделі об'єкта.

Перевіримо спостереження представленої системи. Розрахуємо ранг матриці спостереження за формулою:

$$rank(Z) = rank(H^T, A^T, H^T, \dots, (A^{n-1})H^T), \quad (14)$$

де Z – матриця спостереження, A – матриця системи, H – матриця спостережень, n – розмірність системи.

В результаті обчислення було отримано, що $rank(Z) = 6$, тобто дорівнює порядку системи, отже згідно з критерієм спостереження, система є спостережуваною.

Для перевірки роботи фільтра було проведено експеримент, у якому на БпЛА із встановленою нею системою управління курсом впливав вітер. Результати експерименту наведено на рис. 2 та 3.

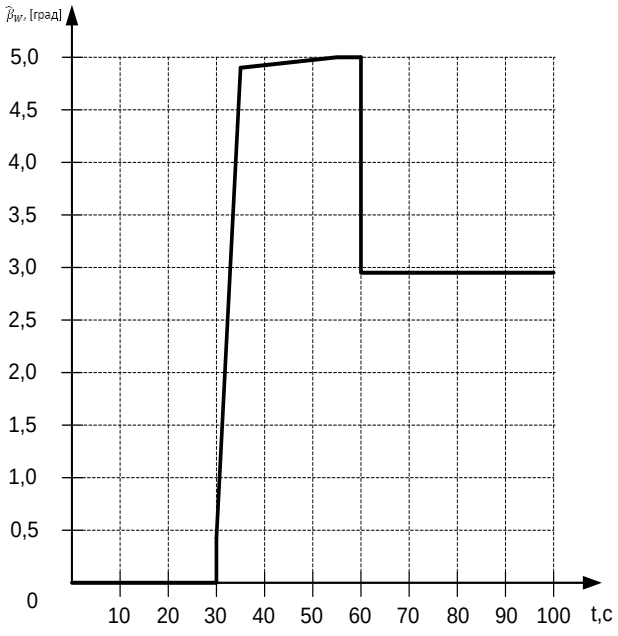


Рис. 2. Зміна кута вітрового зносу БпЛА

Рис. 3. Кут ковзання БпЛА та його оцінка

Як видно з результатів експерименту, значення кута ковзання було оцінено з незначною малою похибкою, а час оцінки не перевищував однієї секунди.

Результати оцінки кута вітрового зносу за аналогічних вітрових впливів наведено на рис. 4.

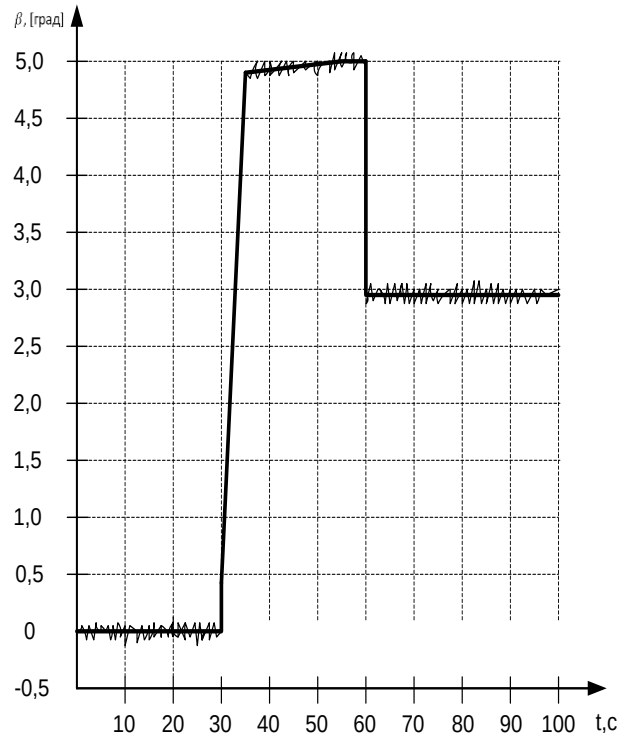


Рис. 4. Результати експерименту з ідеальною математичною моделлю БпЛА (реальне значення кута вітрового зносу та його оцінка)

Висновки

Як видно з наведених графіків, даний метод досить точно оцінює кут вітрового зносу навіть при відхиленні параметрів математичної моделі й може бути застосований при компенсації впливу бокового вітру на безпілотного літального апарату.

Таким чином, замість нестаціонарного оптимального ФК отримано стаціонарний ідентифікатор, що може ефективно оцінювати бічний вітер та зменшувати ступінь завантаження бортового обчислювального пристрою в процесі виконання маневрів при моніторингу елементів критичної інфраструктури.

СПИСОК ЛІТЕРАТУРИ

1. Закон України “Про внесення змін до деяких законів України щодо повноважень уповноваженого органу у сфері захисту критичної інфраструктури України” 18 жовтня 2022 року № 2684-IX. Режим доступу https://zakon.rada.gov.ua/laws/show/2684-20#Text
2. Шмигаль Д.В. За час повномасштабної війни уражено понад 400 об'єктів критичної інфраструктури у сфері електро- і теплопостачання. Електронний ресурс. Режим доступу https://interfax.com.ua/news/general/871517.html
3. Bae, K.-Y., Kim, Y.-D., & Han, J.-H. (2015). Finding a risk-constrained shortest path for an unmanned combat vehicle. Computers & Industrial Engineering, 80, 245–253. doi:10.1016/j.cie.2014.12.016
4. Shreyamsh Kamate, Nuri Yilmazer. Application of Object Detectionand Tracking Techniques for Unmanned Aerial Vehicles. Procedia Computer Science. 2015. № 61. P. 436–441. doi.org/10.1016/j.procs.2015.09.183
5. Kharchenko, V., Sachenko, A., Kochan, V., Fesenko, H. (2016), “Reliability and Survivability Models of Integrated Drone-Based Systems for Post Emergency Monitoring of NPPs”, Proceeding of The International Conference on Information and Digital Technologies 2016, IDT 2016, July 5–7, 2016, Rzeszow, Poland, pp. 127–132 (IEEE Catalog Number CFP16CDT-USB). doi**:** 10.1109/DT.2016.7557161
6. Математичні методи дослідження операцій. Лінейне програмування. Частина 1: навч. пос. / А. А. Яровий, Л. М. Ваховська, Л. В. Крилик. – Вінниця : ВНТУ, 2020. – 86 с. Електронний ресурс. Режим доступу http://ir.lib.vntu.edu.ua//handle/123456789/30859
7. Current Trends in Communication and Information Technologies. Mykhailo Ilchenko, Petro Vorobiyenko, Iryna Strelkovska.IPF 2020, volume 212, Number of Pages XIX, 438. Режим доступу https://doi.org/10.1007/978-3-030-76343-5
8. Глотов В. М., Фис М. М. Застосування БПЛА у військовій справі та аерозніманні. Львів. Львівська політехніка, 2022, 196 с. SBN: 978-966-941-700-8.
9. Ait Saadi, A.; Soukane, A.; Meraihi, Y.; Benmessaoud Gabis, A.; Mirjalili, S.; Ramdane-Cherif, A. UAV path planning using optimization approaches: A survey. Arch. Comput. Methods Eng. 2022, 29, 4233–4284. . Електронний ресурс. Режим доступу https://doi.org/10.1007/s11831-022-09742-7
10. Bal, M. An overview of path planning technologies for unmanned aerial vehicles. Therm. Sci. 2022, 26, 2865–2876. . Електронний ресурс. Режим доступу https://doi.org/10.2298/TSCI2204865B
11. Wu, Y. A survey on population-based meta-heuristic algorithms for motion planning of aircraft. Swarm Evol. Comput. 2021, 62, 100844. . Електронний ресурс. Режим доступу https://doi.org/10.1016/j.swevo.2021.100844
12. Majeed, A.; Hwang, S.O. A multi-objective coverage path planning algorithm for UAVs to cover spatially distributed regions in urban environments. Aerospace 2021, 8, 343. . Електронний ресурс. Режим доступу https://doi.org/10.3390/aerospace8110343
13. Saeed, R.A.; Omri, M.; Abdel-Khalek, S.; Ali, E.S.; Alotaibi, M.F. Optimal path planning for drones based on swarm intelligence algorithm. Neural Comput. Appl. 2022, 34, 10133–10155. Електронний ресурс. Режим доступу https://doi.org/10.1007/s00521-022-06998-9
14. Moriya, N. (2011). Primer to Kalman Filtering: A Physicist Perspective. New York: Nova Science Publishers, Inc ISBN 978-1-61668-311-5. Електронний ресурс. Режим доступу https://www.scribd.com/document/660055294/Kalman-filter
15. Ситник О. О. Аналіз алгоритмів оптимальної фільтрації за показниками точності при скалярних вимірюваннях / Раєвський М. В., Кисельова Г. О. // Відбір і обробка інформації. – Київ : Наукова думка, 2011. – Вип. 34 (110). – ISSN 0474-8662

Received (Надійшла) 29.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

The model of crosswind accounting for flight route planning of unmanned aerial vehicle

Ruslan Kulish

Abstract. In the article, the model of crosswind accounting for flight route planning of unmanned aerial vehicle (UAV) has been developed, which is based on the construction of the optimal Kalman filter. This model is based on modern methods of the flight path optimizing of an unmanned aerial vehicle (UAV), taking into account the existing limitations. The model allows adjusting the waypoints of UAVs, performing monitoring tasks of critical infrastructure elements, which reduces the flight route of UAVs while performing monitoring tasks with taking into account adverse flight conditions and it reduces the loading degree of the UAVs onboard computing device.

Keywords: unmanned aerial vehicle, routing, monitoring of critical infrastructure objects, route planning.

UDC 623.4.011 **doi: 10.26906/SUNZ.2024.1.021**

Elshan Hashimov[1,2], Elnur Khudeynatov[1]

[1] National Defense University, Baku, Azerbaijan
[2] Azerbaijan Technical University, Baku, Azerbaijan

METHODOLOGY FOR ASSESSING THE EFFECTIVENESS OF THE AIR DEFENSE SYSTEM

Abstract. This academic article introduces a pioneering methodology for the comprehensive assessment of air defense systems, addressing existing shortcomings in evaluation approaches. The **subject of this study** revolves around the development and implementation of an advanced methodology for assessing air defense systems. e**The primary aim is** to rectify existing evaluation shortcomings by introducing a holistic model that significantly enhances both efficiency and effectiveness in the evaluation of air defense systems. This methodology incorporates critical elements such as system description, mission, objectives, combat environment, threat, and concept. By explicitly considering factors like system flexibility, survivability, and operational concepts, the study surpasses traditional evaluations, providing a nuanced understanding of the capabilities inherent in air defense systems. The methodology places particular emphasis on the importance of threat analysis, addressing uncertainties related to enemy forces and tactics. Additionally, the study introduces an innovative evaluation model employing tactical scenarios to assess system reliability, availability, and durability, integrating combat environment factors and potential adversary combat options. The research contributes to the academic discourse by providing a systematic and thorough approach to air defense system evaluation, tailored to the complexities of modern warfare and in alignment with the evolving military and technological landscapes. The article suggests avenues for future research to delve into nuanced criteria selection and aggregation methods, aiming to further refine the proposed evaluation methodology.

Keywords: air defense, system, effectiveness, evaluation, concept.

Introduction

Air defense systems play a pivotal role in safeguarding airspace and critical assets from an array of aerial threats including aircraft, missiles, drones, and other airborne platforms. Despite their significant evolution, these systems encounter persistent difficulties, limitations, and challenges [1]. Air defense systems demand meticulous design and management, necessitating a broad operational scope, integration of cutting-edge technologies, robust cybersecurity protocols, and ample resources to sustain continual readiness and effectiveness. This has been a consistent concern in recent literature [2]. Strategic air defense planning, accompanied by effective training programs and sound management practices, becomes imperative to mitigate vulnerabilities commonly associated with air defense systems.

Before delving into the concept of air defense system effectiveness, it is essential to elucidate the distinctions between efficiency and effectiveness. As per P. Drucker's insights, efficiency denotes "doing things right," while effectiveness signifies "doing the right things" [3]. Efficiency is quantified by the ratio of effects to costs, reflecting the precision in execution, whereas effectiveness gauges the extent to which objectives are accomplished.

Literature review. In this context, the effectiveness of an air defense system is gauged by its capability to fulfill predefined objectives and tasks within specific temporal and operational constraints. A myriad of methods exists for evaluating air defense system effectiveness, encompassing metrics like average fire effectiveness, anti-aircraft defense effectiveness, combat activities effectiveness, interaction effectiveness, and airspace control system effectiveness found in technical and tactical literature.

One prevalent calculation method involves determining the efficiency index (E_{AD}) by comparing the average number of destroyed air targets (M_{AD}) to the expected number of targets (N_{EAV}) through the formula [4]:

$$E_{AD} = \frac{M_{AD}}{N_{EAV}} * 100\%, \quad (1)$$

where, E_{AD} is the efficiency index, M_{AD} represents air defense capability, and N_{EAV} is the number of air attack vehicles affecting protected objects.

Another method, endorsed by the US Weapons System Effectiveness Industry Advisory Committee, introduces the effectiveness as a product of availability (A), reliability (R), and capability (C), depicted by the formula [5]:

$$Effektiveness = A * R * C. \quad (2)$$

Here, availability hinges on the system's initial state, reliability mirrors the system's state during task execution, and capability outlines the system's adeptness to perform tasks under specified conditions.

Analyzing these methodologies reveals gaps in accounting for crucial air defense system features. The oversight in selecting the appropriate system, defining accurate criteria, or inadequately describing the operational conditions underscores the necessity for a systematic approach to address these aspects. The ensuing methodology is conceived in response to these considerations, aiming to offer a methodical and logical framework for a comprehensive assessment.

The analysis of these methodologies shows that some important features related to the air defense system are not taken into account during such assessment. Failure to select the appropriate system, correct criteria, or incomplete description of the conditions surrounding air defense system operations indicates the need for a systematic means of determining such aspects. The methodology proposed below arises from the need to pay attention to these considerations and is intended to provide a logical approach to them.

A new methodology for evaluating the effectiveness of the air defense system

A novel methodology is proposed to assess the effectiveness of air defense systems, utilizing system effectiveness as a criterion for comparing alternative systems. The efficacy of a military system is often measured by the degree of success anticipated in achieving its goals. However, as system effectiveness is inherently challenging to gauge qualitatively, the need arises for a quantitative expression to precisely define it. This necessitates the development of a model that encapsulates the pivotal elements influencing the system's effectiveness.

In the realm of air defense systems, critical key elements encompass "description," "mission," "target," "environment," "threat," and "concept." The clarity in defining the system for evaluation is of paramount importance, albeit occasionally challenging due to dependencies on subsystems, neighboring components, and parent systems. For instance, in evaluating an Air Defense division, the primary system may be the AD brigade, with its batteries serving as subsystems.

Systems are intricately designed to furnish interoperable functionality and often function as sub-components within larger systems. Anti-aircraft missile units, strategically positioned to ensure fire system integrity in case of unit failure, establish detection zones at varying heights, exchange vital information, and safeguard the system until readiness for subsequent firing in combat operations. Consequently, the imperative to meticulously define the system for evaluation arises, steering clear of narrow subsystem descriptions that may lead to optimization challenges or overarching parent system descriptions that could yield misleading or impractical analyses.

The evaluation process mandates a clear definition of the system under scrutiny, accompanied by a comprehensive description of its mechanics, functionality, and operational procedures. Understanding the intricacies of how the system mechanically functions, serves its purpose, and is operated forms the foundation for a rigorous evaluation.

System description

Generally speaking, a system description is a section of a technical document or report that provides an overview of the system, its structure and components, and explains how it works [6]. It may also provide information about related systems and technologies used in conjunction with the main system. The initial phase of evaluating an air defense system involves identifying and delineating the system at the appropriate level. The selection of the system for evaluation necessitates careful consideration of whether the air defense means are deployed in localized conflicts or extensive air warfare scenarios directed towards strategic objectives.

Once the system is clearly defined and chosen, the subsequent step involves a detailed description, as exemplified in Fig. 1 for an air defense system. Factors such as vulnerability, resilience, and survival, while challenging to quantify precisely, play pivotal roles in articulating the system's characteristics. The vulnerability of the air defense system emerges as a critical element, given that enemy operations typically involve penetrating or neutralizing air defenses, necessitating countermeasures against electronic interference, damage, or destruction. Therefore, the system's vulnerability to electronic countermeasures, high-precision weaponry, anti-radar missiles, air-to-ground missiles, and conventional ammunition must be consistently considered.

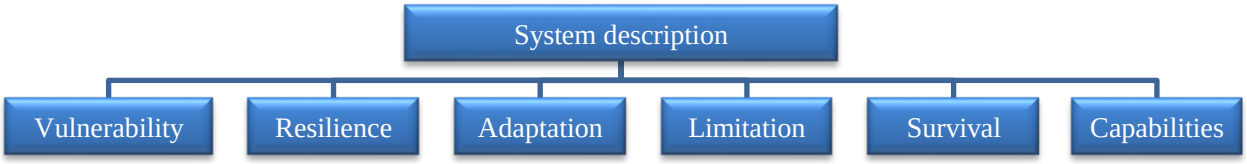


Fig. 1. System description

System flexibility is an integral aspect of its description, acknowledging that changes in tasks, locations, or operational modes have cascading effects on other factors, including capabilities and limitations.

Survivability, though subtle, constitutes a crucial component of the system description. The success or failure of a firing unit within the broader air defense system can exert a direct influence on the overall system's functionality. Key elements in system description encompass capabilities, limitation, and compatibility. In essence, these factors delineate the system's operational scope, limitations, and its integration into overall combat operations. While combat capability stands out as a quantifiable and information-rich indicator, a comprehensive understanding requires simultaneous consideration of limitation and compatibility. System limitations, encompassing mission completion time, reload time, weather conditions, and combat resources, introduce measurability and realism into the system description. Although quantifying compatibility is more intricate than capabilities and limitation, it is a crucial factor defining the air defense system's ability to interact with ground and air combat assets effectively, emphasizing the need for efficient management, control,

and information exchange interfaces for individual firing units to fulfill assigned tasks.

Mission

A mission describes what the system will do and the purpose of doing it. The mission statement describes Kipling's "six honest serving-men" – who, what, when, where, why, and sometimes how. The mission provides the context for defining measures of effectiveness and for development of the Concept of Operations [7]. The mission of an air defense system constitutes a pivotal element in its effectiveness assessment, necessitating the translation of the system's mission into clearly defined objectives.

The complexity of multitasking can lead to an excessive number of system goals, creating challenges in decision-making processes. To discern the system's mission, it is imperative to initially establish a comprehensive concept of the system. Concept definition involves a meticulous examination of the needs and requirements within the problem area, incorporating general processes such as task analysis and consideration of stakeholder needs and requirements [8].

Task analysis, a method to distinctly identify the problem and available means to resolve it, encompasses activities such as delineating the system's goals and formulating the operational concept.

The determination of system objectives plays a crucial role in enhancing its effectiveness, demanding an analytical approach based on the best available information about warfare. The clarity and precision with which the mission is expressed directly influence the logical transformation of requirements into desired results. In the context of air defense systems, considering both air and surface combat requirements is essential to avoid the complications associated with multitasking.

Objectives, categorized in Fig. 2, are derived from various sources, including doctrine and considerations specific to the system's objectives and constraints. Doctrinal statements, such as those ensuring air defense within a unified AD system, provide a general framework applicable to air defense assessments [9]. Implicit and explicit requirements guide these objectives, ranging from explicit goals concerning enemy assets to veiled demands emerging from tactical shifts, such as terrain masking during deployment.

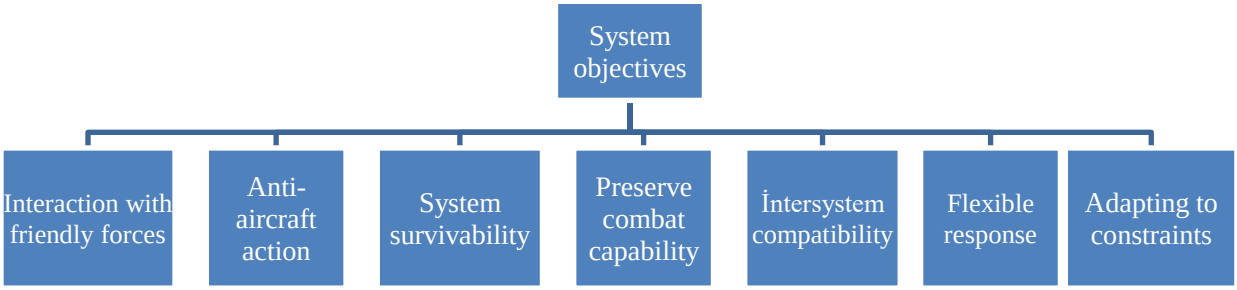


Fig. 2. Objectives of the system

Continuous combat readiness, a primary goal of system evaluation, comprises elements like system reliability, availability, and durability. The detailed analysis of combat-ready, weakened, and non-combat-ready firing units is more informative than a single figure reflecting the total number of assets.

The articulation of clearly stated objectives, coupled with a thorough definition and description of the system, establishes the foundational concept of effectiveness for evaluation. For instance, an air defense system tasked with protecting forces from enemy air attacks would evaluate its qualitative effectiveness based on the degree to which it safeguards forces and key facilities within its area of responsibility.

Considering the objectives of an air defense system underscores the requirement for its means to support both air and ground combat. The task of defending the homeland against air threats serves as a clear and precise organizational purpose, guiding the entire system development process. This requirement acts as a reference point for determining the suitability of the system's performance in fulfilling its mission.

Concept of operation

A concept of operation is a document that describes a proposed system concept and how that concept would be operated in an intended environment [10]. The operational concept serves as a verbal and graphical representation of an organization's assumptions or intent regarding the operation of a system or a collection of related systems [11].

This concept aims to provide an overview of operations from the users' perspective, offering insights into the use of specific systems within an organization's operating environment.

In the context of the Air Defense System, the overarching goal is to defend the homeland by neutralizing the unwanted effects of enemy air attack means.

The operational concept can be visually represented in Fig. 3, where broken lines connect external objects that, while not direct components of the system, play crucial roles in its operation.

External Objects:

Air Threats: Encompassing the enemy's air and missile capabilities, this external object is pivotal in determining system requirements and directly influences task accomplishment.

Weather and Visibility Conditions: Despite sensitivity to weather, the air defense system must ensure continuous combat readiness, considering diverse weather conditions.

Operational Directives: Set at the political level, these directives dictate the capabilities required for the Air Defense System to fulfill its mission aligned with the country's political interests.

Naval Air Defense Assets: Though not direct components, these tools enhance the system's detection and destruction capabilities.

Internal Objects:

C4ISR: Concentrating command, control, management, surveillance, and intelligence functions, these tools facilitate effective coordination and decision-making among system objects.

Radars: Emphasizing the key capability of detection, these tools are crucial for tracking and neutralizing potential threats, ensuring the system fulfills its mission.

Fighter Aviation: Providing essential information about detected unknown objects, fighter aviation determines the threat level and can engage in combat activities against air targets.

SAM Systems: Highly effective in destroying both manned and unmanned aerial vehicles, including cruise missiles, these systems require strategic repositioning and have limitations in addressing unknown targets.

Air Defense Units of the Troops: Equipped with anti-aircraft artillery and portable missiles, these units protect troops from direct air threats.

The operational concept unfolds as follows: the initial step involves detecting potential air threats under any weather or visibility conditions. Upon detection, command posts are activated, referencing operational directives to counter the threat effectively. If the identified target is confirmed as an enemy, it can be neutralized using anti-aircraft missiles or artillery. If additional information is needed for target identification, fighter jets are deployed, and if necessary, the target is destroyed.

This systematic approach ensures a comprehensive response to potential air threats, aligning with the mission of defending the homeland.

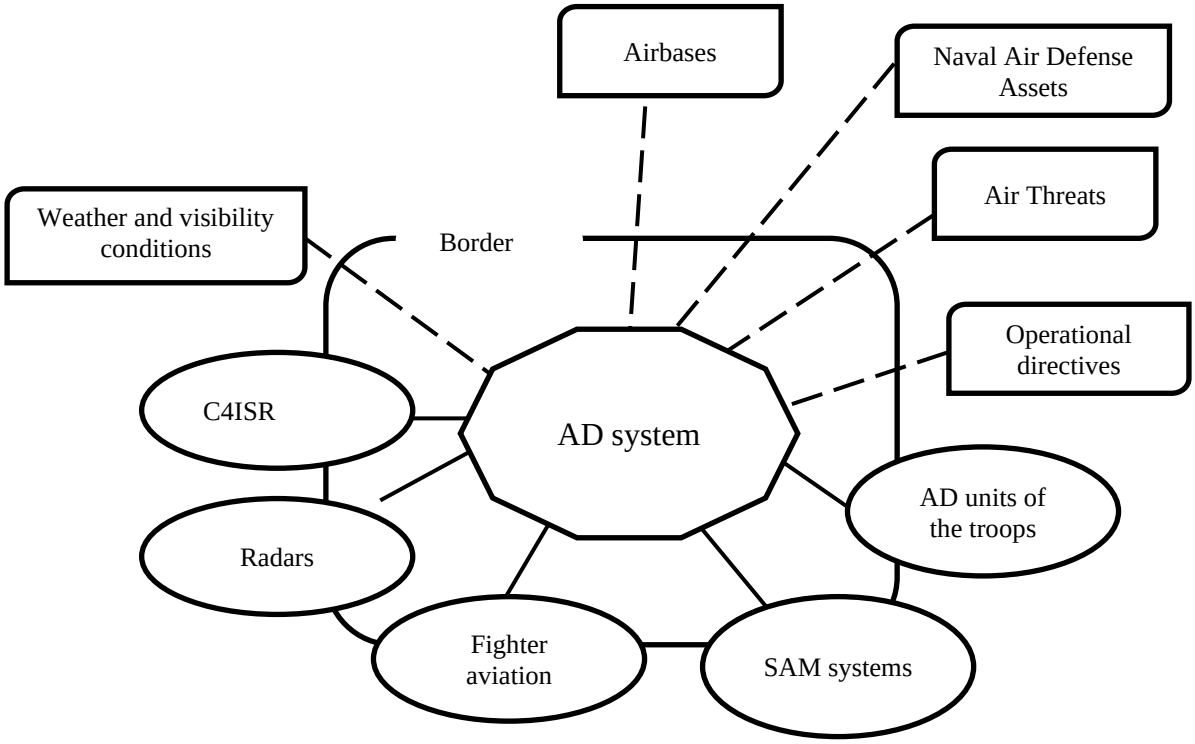


Fig. 3. Operational concept of AD system

Threat

In evaluating system effectiveness, a precise articulation of threats is pivotal. Threats are elements surrounded by uncertainty, grounded in intelligence but necessitating judgment regarding feasibility. Clear threat definition involves considering the enemy's target forces and tactics, forming the primary components of the tactical scenario essential for assessment (Figure 4).

The air defense system utilizes a threat tactical scenario to devise countermeasures, assessing its capability to safeguard protected objects. While multiple elements contribute to threat development, they can generally be classified within the components of Fig. 4.

Target Objectives:

Determining where and how to neutralize air defenses or navigate strong defense points through indirect approaches, assessing the enemy's objectives involves acquiring sufficient information. Factors such as target priority, desired damage level, maximum acceptable attrition, and the enemy's influence on operations are crucial for accurate assessment.

Force Data:

Dependent on the selected system's rating level, force data encompasses the enemy's combat capability, battle formation, and organization-staff structure. This information is vital for anticipating possible threat situations.

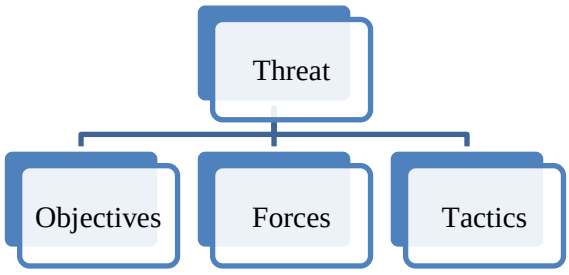


Fig. 4. Threat factors

Tactics:

Tactics play a significant role in defense assessment against air attacks, particularly concerning protected objects. Understanding the enemy's tactics and their rules of application is essential. War games can aid in determining which combat methods and tactics merit consideration.

Threat clarification for assessment combines objective, force, and tactics, translating into the enemy's flight profile, attack method, and approach directions. These elements are integrated into assessment models, transforming information into a series of combat options that the system must contend with, judge, and quantify. This process contributes to developing an operational definition of system effectiveness.

For an air defense system designed to protect objects, the enemy's intent might be air bases, headquarters, etc.

The effectiveness of the air defense system is contingent on quantifying information such as approach paths, flight profiles, aircraft types, and target priorities. When incorporated into a tactical scenario, the enemy's battle tactics emerge, providing a comprehensive evaluation of the Air Defense system.

Combat environment

In the evaluation of air defense system effectiveness, a crucial requirement is the comprehensive analysis of the combat environment. This assessment is pivotal for two primary reasons. Firstly, understanding the combat environment, coupled with the system description, enables the determination of the combat readiness state.

The environment, with its components such as weather, jamming, and terrain, exerts a profound influence on the system's condition. Secondly, the combat environment, alongside force and task considerations, serves as a fundamental element in shaping the tactical scenario that underpins the assessment process. Numerous elements contribute to the combat environment, logically falling into categories depicted in Figure 5. While military-economic factors are typically established at a level beyond direct assessment, their impact can be substantial in the evaluation.

As the focus narrows on the combat environment, weather and terrain emerge as primary considerations due to their significant effects on camouflage, mobility, and the performance degradation of systems in adverse weather conditions. Evaluating electronic countermeasures (ECM) and anti-radiolocation missiles (ARM) necessitates distinguishing whether they originate from air or surface means.

In a scenario where adversaries capable of mounting effective air attacks are present, it is implausible for them to risk expensive and modern aircraft in an environment lacking tactical efficiency.

Therefore, if Air Defense systems are engaged in combat, it must be assumed that the adversary will employ electronic countermeasures and anti-radar missiles extensively. The evaluation of the Air Defense system must align with this assumption, intricately incorporating it into the development of the combat environment.

Evaluation process

In evaluating the effectiveness of an air defense system, a robust evaluation model is essential, employing a tactical scenario to operationally ascertain the system's capabilities. This scenario delineates what should be observed, the conditions under which observations occur, and the requisite operations. The assessment of system effectiveness must encompass how observations are made, measured, and managed. By amalgamating the results of the assessed tactical scenario with selected effectiveness criteria, a comprehensive evaluation of the system's effectiveness is achieved. The tactical scenario results can be derived through parametric analysis, system indices, deterministic or stochastic models, as well as analyst or user judgment.

Fig. 6 visually outlines the integral elements of the assessment process, where the system description, operational concept, and combat environment collectively shape potential combat readiness situations.

The combat environment, coupled with threat information, generates a spectrum of adversary combat options for system testing. Parameters like reliability, availability, and durability are derived from states of combat readiness, influenced by the enemy's warfare methods to determine the system's reliability.

As an illustrative example, consider a combat operation involving air defense, ground forces, and fighter aircraft countering enemy air strikes. The tactical scenario unfolds with ground troops in a defensive zone, air defenses in combat positions, and fighter aircraft ready on the runway. The objective is to protect objects, while the enemy aims to destroy selected targets using a low flight profile.

The combat environment is simplified (no electronic countermeasures/anti-radar missile application). Assuming all systems are combat-ready, the tactical scenario progresses, evaluating combat operations using various assessment models. These models assess target detection, tracking, identification, decoy, and other elements. The evaluation includes reports detailing subsystems neutralized or destroyed, surviving assets, system damage, enemy air attack assets' status, and other criteria describing combat operations.

Returning to the example, if the enemy shifts tactics due to pressure from air defenses, achieving medium-altitude profiles, and coordination with fighter aircraft results in successful interaction, the air defense system's effectiveness is demonstrated.

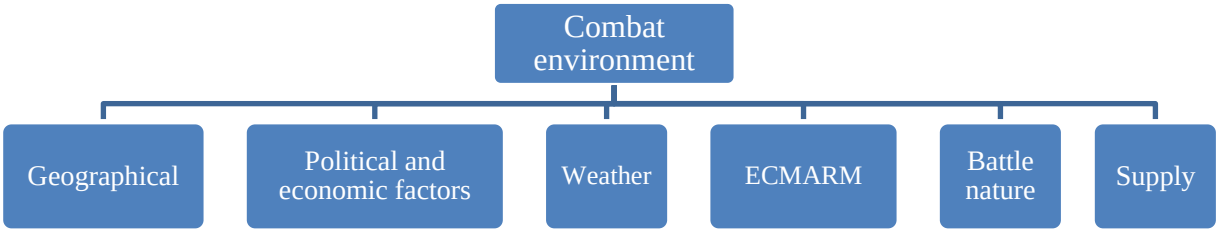


Fig. 5. Combat environment

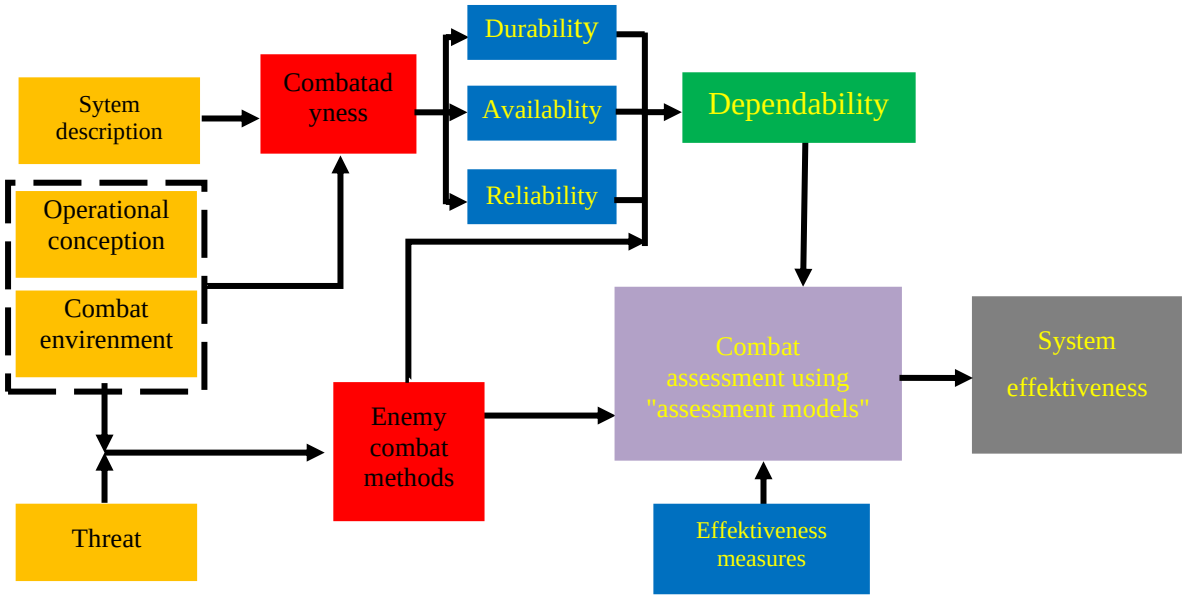


Fig.6. Evaluation model

The combat assessment yields quantitative information supporting the system's evaluation objectives.

Effectiveness criteria selection is pivotal in the evaluation, reflecting the degree of goal achievement. Choosing measures that accurately depict the system's effectiveness is particularly crucial in evaluating air defense systems, considering the potential for multiple effectiveness criteria.

Therefore, a careful selection of measures is imperative to accurately gauge the evaluated system's effectiveness accurately depict the system's effectiveness is particularly crucial in evaluating air defense systems, considering the potential for multiple effectiveness criteria. Therefore, a careful selection of measures is imperative to accurately gauge the evaluated system's effectiveness.

Combat assessment can include a number of assessment models that assess target detection, tracking, identification, decoy, and other elements of combat operations [13-15]. These models might be Monte Carlo Simulation Models, Mathematical Models for Sensor Performance, Recognition Models, Electronic Warfare Models, Battlefield Simulation Models, Operational Analysis Models, C4ISR Models [16-18].

Conclusion

In the contemporary landscape of air defense systems, measuring effectiveness solely through the destruction of the enemy's air attack means for mission accomplishment is deemed inadequate. Alternative approaches include dissuading the enemy from executing the task, compelling the utilization of unfavorable attack profiles, or prompting fewer impact strikes.

Another avenue for establishing performance criteria involves summarizing or weighting them based on their significance.

While a single criterion poses no challenge, the presence of multiple efficiency criteria necessitates determining the most crucial ones for gauging system effectiveness. This decision, often made by assessors, can benefit from expert opinions, consultation with decision-makers on prioritized actions, or the application of a decision matrix.

In determining the overall system effectiveness, it is prudent to devise an efficiency function offering a measure of system effectiveness for comparison among alternatives.

Formula 2 could yield the desired outcome if the system is defined as a unit of fire. However, if the system encompasses an Air Defense combination or operates at the operational level, assessing effectiveness involves considering the defense in the combat environment, the adversary, and their interactions through comprehensive effectiveness criteria.

Evaluations at general levels tend to be intricate, requiring the careful selection, aggregation, and combination of key performance criteria with combat evaluation data to furnish a reliable indication of system effectiveness.

REFERENCES

1. Hashimov, E.G., Khudeynatov, E.K. V-Model for Air Defense Systems // Modeling, control and information Technologies: Proceedings of VI İnternational scientific and practical conference. No.6 (2023). –Rivne: november 9-11, -2023, -p.46-49. https://doi.org/10.31713/MCIT.2023.011
2. Hashimov, E.G., Khudeynatov, E.K. The effectiveness of air defense system // Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління. Тези доповідей тринадцятої міжнародної науково-технічної конференції, - Харків: - 26 – 27 квітня, -2023, Том 1. - pp.17-18. https://doi.org/10.32620/%20ICT.23.t1 URL:https://nure.ua/wp-content/uploads/2023/tom_1_ict_2023.pdf#page=4
3. Staff Writer. Effectiveness vs. efficiency: What you need to achieve both. [Electronic resource] / Smartsheet, available at: https://www.smartsheet.com/content-center/managing-work/effectiveness-vs-efficiency-what-you-need-achieve-both
4. Daniel Michalski, Adam Radomyski. Counting the Uncountable: Introduction to the New Method of Evaluation of the Efficiency of Air Defense // Safety and defense, 2020, Vol 6, №2. P.100-112. Doi:https://doi.org/10.37105/sd.91
5. Monroe, Alfred J., Voegtlen, H. D. , Moxley Jr, Frank H. Weapon System Effectiveness Industry Advisory Committee (WSEIAC). Final Report of Task Group VI, Chairman's Final Report (Integrated Summary). [Electronic resource] / Defense Technical İnformation Center, available at: https://apps.dtic.mil/sti/citations/AD0467816
6. System Description (Section III). [Electronic resource] / Scytale, available at: https://scytale.ai/glossary/system-description-section-iii/
7. Giachetti R., Hernandez A., Mission Engineering [Electronic resource] / SEBoK, available at: https://sebokwiki.org/wiki/Mission_Engineering
8. Roedler G., Adcock R., Business and Mission Analysis. [Electronic resource] / SEBoK, available at: https://sebokwiki.org/wiki/Business_and_Mission_Analysis
9. Military Decisionmaking Process. Lessons and best practices // Handbook, No.15-06 p.11, available at: https://usacac.army.mil/sites/default/files/publications/15-06_0.pdf
10. Systems Engineering. Concept of Operations (CONOPS). [Electronic resource] / AcqNotes. The Defense Acquisition Encyclopedia, available at: https://acqnotes.com/acqnote/careerfields/concept-of-operations-conopsse
11. Military doctrine of the Republic of Azerbaijan. [Electronic resource] / available at: http://anl.az/down/meqale/azerbaycan/2010/iyun/124735.htm
12. Operational Concept (glossary). [Electronic resource] / SEBoK, available at: https://sebokwiki.org/wiki/Operational_Concept_(glossary)
13. Bayramov A. A., Hashimov E. G., Amanov R. R. The detection of invisible objects on the terrain on the basis of GIS technology //Geography and nature sources. – 2016. – C. 124-126.
14. Bayramov A.A., Hashimov E.G. Seismic Location Station for Detection of Unobserved Moving Military Machineries // Journal of Management and Information Science, 2016, Vol.4, № 2, p. 61-66. DOI:10.17858/jmisci.82365
15. Hashimov E. G., Huseynov B. S. Some aspects of the combat capabilities and application of modern UAVs //Baku: "National Security and military knowledges. – 2021. – №. 3. – p. 14-24. URL: http://mmu.edu.az/assets/files/magazine/6c76692129d7fc24.pdf
16. Piriyev H.K., Hashimov E.G., Bayramov A.A. Modelling of the battle operations // Monografiya, Herbi Nashriat", Baku. – 2017. -256 p.
17. He Zhengqiu, Wang Lihua, Liu Wenfu, Xu Zhongfu. Model and Effectiveness Analysis for C4ISR System Structure Based on Complex Network // 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Vol. 01, 85-88 - August 2018. https://doi.org/10.1109/ihmsc.2018.00027
18. Petrovski A., Radovanović M., Behlic A., Ackovska S. Advantages of implementation of C6ISR in low budget armies // International Scientific Conference Geobalcanica. – Skopje: 08-9 may 2023.- p.47-60. DOI: https://doi.org/10.18509/GBP23047p

Received (Надійшла) 08.11.2023
Accepted for publication (Прийнята до друку) 17.01.2024

Методика оцінки ефективності системи протиповітряної оборони

Ельшан Гашимов, Ельнур Худейнатов

Анотація. Ця наукова стаття представляє новаторську методологію всебічної оцінки систем протиповітряної оборони, усуваючи наявні недоліки в підходах до оцінки. **Предметом цього дослідження** є розробка та впровадження вдосконаленої методології оцінки систем протиповітряної оборони. **Основною метою** є усунення існуючих недоліків оцінки шляхом впровадження цілісної моделі, яка значно підвищує як ефективність, так і результативність оцінки систем протиповітряної оборони. Ця **методологія** включає важливі елементи, такі як опис системи, місія, цілі, бойове середовище, загроза та концепція. Завдяки чіткому розгляду таких факторів, як гнучкість системи, живучість і експлуатаційні концепції, дослідження перевершує традиційні оцінки, забезпечуючи детальне розуміння можливостей, притаманних системам ППО. Методологія приділяє особливу увагу важливості аналізу загроз, розглядаючи невизначеності, пов'язані з силами та тактикою противника. Крім того, у дослідженні представлена інноваційна модель оцінки, яка використовує тактичні сценарії для оцінки надійності, доступності та довговічності системи, інтегруючи фактори бойового середовища та потенційні бойові варіанти супротивника. **Дослідження сприяє** науковому дискурсу, забезпечуючи систематичний і ретельний підхід до оцінки системи протиповітряної оборони, пристосований до складнощів сучасної війни та у відповідності до еволюції військових і технологічних ландшафтів. У статті **пропонуються шляхи для майбутніх досліджень**, спрямованих на вивчення нюансів вибору критеріїв і методів агрегування, з метою подальшого вдосконалення запропонованої методології оцінювання.

Ключові слова: ППО, система, ефективність, оцінка, концепція.

УДК 681.7.068.4 **doi: 10.26906/SUNZ.2024.1.028**

О. М. Клименко

Харківський національний університет радіоелектроніки, Харків, Україна

КОНЦЕПЦІЯ РОБОТИ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ НА БАЗІ МЕТОДУ НАЙБЛИЖЧИХ СУСІДІВ ДЛЯ БАЛАНСУВАННЯ НАВАНТАЖЕННЯ НА КОНВЕЄРНИХ ЛІНІЯХ СОРТУВАННЯ ПРОДУКЦІЇ

Анотація. Ця стаття представляє дослідження концепції роботи алгоритмів машинного навчання з учителем на основі методу найближчих сусідів для оптимізації процесу балансування навантаження на фармацевтичних лініях сортування. Сфера фармацевтики стикається з навантаженням, що постійно змінюється, і різноманітними характеристиками упаковок, що вимагає ефективних стратегій управління виробничими процесами. Автор пропонує концепцію, що базується на принципі найближчих сусідів, для визначення оптимального навантаження для кожної зони сортування. Використання цього методу дозволяє враховувати контекст і схожість зон сортування, що призводить до точних та адаптивних рішень. Роботи-сортирувальники, що працюють на фармацевтичних конвеєрних лініях, стають більш ефективними у розподілі упаковок по всіх робочих зонах, що призводить до зниження часу складання замовлень та оптимізації ресурсів.

Ключові слова: індустрія 4.0, розумне виробництво, логістика 4.0, складське господарство 4.0, Інтернет речей, бережливе виробництво, балансування навантаження, сортувальня конвеєрна лінія, фармацевтичне виробництво.

Вступ

У сучасному світі, де концепції Індустрія 4.0, Розумне виробництво, Логістика 4.0, Складське господарство 4.0, відіграють ключову роль у перетворенні виробничих процесів, увага до оптимізації та ефективності у сфері фармацевтики стає критично важливою [1,2]. Науковці розглядають цілу низку проблем, пов'язаних із впровадженням принципів концепції Складське господарство 4.0 (Warehousing 4.0). Серед цих проблем можна виділити використання шатлів [3-5], питання логістики на фармацевтичних підприємствах [6, 7], розробки програмного забезпечення для складів [8], розробки та роботи конвеєрної стрічки [9], використання роботів [10, 12] тощо.

На тлі цього контексту дана стаття спрямована на дослідження застосування передових технологій у галузі балансування навантаження на фармацевтичних лініях сортування, що є надзвичайно актуальною задачею [13-17].

Основним фокусом дослідження є використання алгоритмів машинного навчання з учителем для оптимізації процесу сортування упаковок. Роботи-сортирувальники, що працюють на фармацевтичних конвеєрних лініях, стикаються з навантаженням, що постійно змінюється, і різними характеристиками упаковок.

Ми доповнили наше дослідження завдяки інтеграції Інтернету речей (IoT) та принципів Бережливого виробництва, забезпечуючи систему реального часу для збору даних про навантаження, швидкість та інші фактори, що впливають на ефективність сортування. Цей комплексний підхід не тільки покращує виробничі показники, а й сприяє більш стійкій та адаптивній фармацевтичній логістиці. Актуальність досліджень проявляється у необхідності мінімізації трудомісткості формування замовлень та підвищення ефективності роботи конвеєрних ліній. Запропоновані алгоритм машинного навчання надають інтелектуальні інструменти для досягнення фармацевтичної індустрії, що динамічно розвивається, де ефективність виробництва надзвичайно важливою для забезпечення конкурентоспроможності.

Загальний принцип роботи алгоритму машинного навчання з учителем

Загальний принцип роботи алгоритму машинного навчання з учителем (Supervised Learning) полягає у використанні розмічених даних для навчання моделі, яка здатна робити передбачення на нових, раніше невідомих даних. Цей тип машинного навчання має на увазі наявність "вчителя" або "експерта", який надає моделі дані та відповідні їм правильні відповіді. Основні кроки, які включає процес навчання алгоритму машинного навчання з учителем:

- підготовка даних. Завантаження та попередня обробка даних, включаючи очищення, масштабування, перетворення ознак та інші операції. Поділ даних на навчальний та тестовий набори для оцінки продуктивності моделі;

- вибір моделі. Вибір архітектури моделі в залежності від характеру задачі (класифікація, регресія і т.д.). Вибір алгоритму, який найкраще підходить для вирішення конкретної задачі.

- навчання моделі. Подача навчальних даних на вхід моделі. Модель "навчається" на основі розмічених прикладів, коригуючи свої ваги або параметри для мінімізації помилки між передбаченими значеннями та реальними відповідями;

- оцінка моделі. Використання тестового набору даних для оцінки продуктивності моделі. Використання метрик оцінки (наприклад, точність, F1 міра, середньоквадратична помилка), щоб визначити, наскільки добре модель справляється з новими даними;

- налаштування гіперпараметрів. При необхідності проведення налаштувань гіперпараметрів для покращення продуктивності моделі.

- прогнозування нових даних. Після успішного навчання модель може використовуватися для передбачення відповідей на нові, раніше невідомі дані.

Прикладами алгоритмів машинного навчання з учителем є лінійна регресія, метод найближчих сусідів, дерева рішень, метод опорних векторів (SVM), ансамблі (наприклад, випадковий ліс) та глибокі нейронні мережі.

Розробка методу машинного навчання з учителем на основі методу найближчих сусідів

Метод найближчих сусідів (k-Nearest Neighbors, k-NN) – це алгоритм класифікації та регресії, заснований на принципі близькості об'єктів у просторі ознак. Розглянемо застосування методу найближчих сусідів балансування навантаження на фармацевтичних лініях сортування.

Визначимо у межах даного дослідження під поняттям простір ознак будемо розуміти математичне поняття, що визначає всі можливі комбінації значень ознак кожного об'єкта. Позначимо його як X і кожен об'єкт представлений у цьому просторі як вектор ознак. Нехай X є n _простір, де n - кількість ознак. Кожен об'єкт представлений вектором ознак:

$$x_i = (x_{i1}, x_{i2}, .., x_{in}), \quad (1)$$

де x_{in} - ознака, яка може бути числовим значенням, категоріальним (фактором) або бінарним.

Внаслідок цього матриця ознак X містить усі об'єкти та їх ознаки. Якщо у нас є m об'єктів та n ознак, то матриця буде розміру $m \times n$.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}. \quad (2)$$

У межах цих досліджень уявімо, що є фармацевтична лінія сортування, і необхідно визначити оптимальне навантаження для кожної зони. Ознаками можуть бути такі параметри: поточне навантаження (x_1), швидкість обробки (x_2), ефективність (x_3) та інші параметри. Виходячи з цього вектор ознак для першої робочої зони роботів-сортувальників можна подавати наступним чином:

$$x_1 = (x_{11}, x_{12}, x_{13}). \quad (3)$$

Простір ознак надає абстрактний, математичний фреймворк для подання та аналізу даних, що є ключовим компонентом багатьох методів машинного навчання.

Розглядаючи завдання балансування навантаження на фармацевтичних лініях сортування, цільовою змінною може бути оптимальне навантаження для кожної зони сортування.

Цільова змінна (y) в математичній моделі є величиною, яку ми прагнемо передбачити або пояснити з використанням ознак з простору ознак. У різних завданнях машинного навчання цільова змінна може бути числовою (у задачах регресії) або категоріальною (у задачах класифікації). Якщо ми розглядаємо завдання балансування навантаження на фармацевтичних лініях сортування, цільовою змінною може бути оптимальне навантаження для кожної зони сортування.

У задачі регресії:

$$y_i \in R. \quad (4)$$

У задачі класифікації:

$$y_i \in \{C_1, C_2, ..., C_k\}, \quad (5)$$

де y_i – цільова змінна для зони i.

Мета полягає в побудові моделі $f(X)$, яка максимально точно передбачає або класифікує цільову змінну y на основі вхідних ознак з простору X.

Наступною дією необхідно визначити відстані в математичних моделях, в контексті методу найближчих сусідів (k-NN) це є мірою віддаленості між двома об'єктами в просторі ознак. Це поняття необхідно для визначення "близькості" об'єктів і, таким чином, для ухвалення рішення у методі найближчих сусідів. Нехай нас є два об'єкти, представлені векторами ознак x_i і x_j. Відстань між ними позначимо як $d(x_i, x_j)$ Виходячи з цього можна використовувати такі типи відстаней:

– евклідова відстань:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2}; \quad (6)$$

– манхеттенська відстань

$$d(x_i, x_j) = \sum_{k=1}^{n} | x_{ik} - x_{jk} |; \quad (7)$$

– косинусна відстань

$$d(x_i, x_j) = 1 - \frac{\sum_{k=1}^{n} x_{ik} \cdot x_{jk}}{\sqrt{\sum_{k=1}^{n} (x_{ik})^2} \cdot \sqrt{\sum_{k=1}^{n} (x_{jk})^2}}. \quad (8)$$

Вибір конкретної метрики залежить від природи даних та характеру завдання. Наприклад, евклідова відстань часто використовується, коли важливо враховувати абсолютні значення ознак. При використанні k-NN для кожного об'єкта обчислюються відстані до всіх інших об'єктів у вибірці (6-8). Потім обираються k найближчих сусідів з урахуванням цих відстаней. Тобто для кожної зони сортування x_i знаходимо k найближчих сусідів за обраною метрикою відстані. Зважуємо їх цільові змінні y_i на основі відстані, наприклад через зворотну відстань:

$$y_i = \left(\frac{1}{d(x_i, x_j)}\right). \quad (9)$$

Це дає можливість обчислити зважене середнє для визначення оптимального навантаження для зони x_i.

Коли з'являються нові дані (нова зона сортування), використовуємо алгоритм k-NN для прогнозування оптимального навантаження на основі найближчих сусідів.

Висновки

У процесі дослідження концепції роботи алгоритмів машинного навчання з учителем, заснованих на методі найближчих сусідів, для балансування навантаження на конвеєрних лініях сортування при розподілі упаковок по робочих зонах роботів сортувальників було виявлено значні та перспективні результати: моделі, засновані на даному методі, успішно прогнозують оптимальні навантаження для різних зон сортування; враховуються їх характеристики та розподіл упаковок.

Використання різних метрик відстані, таких як евклідова, манхеттенська та косинусна відстані, дозволило обрати найкращу метрику в залежності від особливостей конкретного завдання. Це додатково підкреслило гнучкість та адаптивність методу до різноманітних умов роботи фармацевтичних ліній сортування.

Отримані результати мають пряме практичне застосування у фармацевтичній індустрії та можуть бути узагальнені на інші галузі промисловості з аналогічними принципами сортування та розподілу продукції.

СПИСОК ЛІТЕРАТУРИ

1. Yevsieiev, V., & Gurin, D. (2023). Comparative Analysis of the Basic Methods Used in Industry 4.0 and Industry 5.0. Collection of Scientific Papers «ΛΌГОΣ», (Bologna, Italy), 113–115. https://doi.org/10.36074/logos-29.09.2023.31
2. Nevliudov, I., & et al. (2020). Method of Algorithms for Cyber-Physical Production Systems Functioning Synthesis. International Journal of Emerging Trends in Engineering Research, 8(10), 7465-7473.
3. Maksymova, S., & et al. (2023). Shuttle-Based Storage And Retrieval System 3d Model Improvement and Development. Journal of Natural Sciences and Technologies, 2(2), 232-237.
4. Nevliudov, I., & et al. (2022). Analysis of Software Products for Simulation Modeling of the Operation of the System of Shuttles for Warehousing. In Manufacturing & Mechatronic Systems 2022: Proceedings of VIst Int. Conf., Kharkiv,24-26.
5. Igor, N., & et al. (2023). (2023). Using Mecanum Wheels for Radio Shuttle. Multidisciplinary Journal of Science and Technology, 3(3), 182-187.
6. Igor, N., & et al. (2023). Automated Logistics Processes Improvement in Logistics Facilities. Multidisciplinary Journal of Science and Technology, 3(3), 157-170.
7. Nevliudov, I., & et al. (2023). Features of Wave Algorithm Application in Warehouse Logistics Transport Systems. Information systems in project and program management: Coll. mon. European University Press. Riga: ISMA, 251-261.
8. Nevliudov, I., & et al. (2023). Software development for small details production warehouse automated system. Scientific Collection «InterConf», 320-323.
9. V. V. Yevsieiev, & et al. (2023) Conveyor Belt Object Identification: Mathematical, Algorithmic, and Software Support. Appl. Math. Inf. Sci. 17, No. 6. - P. 1073-1088.
10. Yevsieiev, V., & Gurin, D. (2023) Comparative Analysis of the Characteristics of Mobile Robots and Collaboration Robots Within INDUSTRY 5.0. In the VI International Scientific and Theoretical Conf., September 8, 2023. Chicago, USA. 92-94
11. Yevsieiev V. Some aspects of the development of the BEAM robot control scheme. In IV International Scientific and Theoretical Conference,Singapore, Republic of Singapore. - P. 79-81.
12. Невлюдов І. Ш. та інші (2024) BEAM робототехніка : навч. посіб. Кривий Ріг : Вид. Чернявський Д. О., 2024. – 276 с.
13. Kuo, Yiyo, Ssu-Han Chen, Taho Yang, and Wei-Chen Hsu. 2023. "Optimizing a U-Shaped Conveyor Assembly Line Balancing Problem Considering Walking Times between Assembly Tasks" Applied Sciences 13, no. 6: 3702. https://doi.org/10.3390/app13063702
14. Zhang, S. & Xia, X. (2010). Optimal control of operation efficiency of belt conveyor systems. Appl. Energy, 87, 1929–1937.
15. Ponnambalam, S.G. & et al. (1999). A comparative evaluation of assembly line balancing heuristics. Int. J. Adv. Manuf. Technol. 1999, 15, 577–586.
16. Oksuz, M.K.; Buyukozkan, K.; Satoglu, S.I. U-shaped assembly line worker assignment and balancing problem: A mathematical model and two meta-heuristics. Comput. Ind. Eng. 2017, 112, 246–263.
17. Khalid, M. S., & et al. (2022). HMI Development Automation with GUI Elements for Object-Oriented Programming Languages Implementation. International Journal of Engineering Trends and Technology, 70.1, 139-145.

Received (Надійшла) 23.11.2023
Accepted for publication (Прийнята до друку) 17.01.2024

Concept of supervised machine learning algorithms based on the k-nearest neighbors method for load balancing on pharmaceutical sorting lines

Oleksandr Klymenko

Abstract. This article presents a study on the concept of supervised machine learning algorithms based on the k-nearest neighbors method to optimize the load balancing process on pharmaceutical sorting lines. The pharmaceutical industry faces constantly changing loads and diverse packaging characteristics, requiring efficient strategies for managing production processes. The author proposes a concept based on the nearest neighbors principle to determine the optimal load for each sorting zone. The use of this method allows considering the context and similarity of sorting zones, leading to accurate and adaptive solutions. Sorting robots operating on pharmaceutical conveyor lines become more efficient in distributing packages across all working zones, resulting in reduced order assembly time and resource optimization.

Keywords: Industry 4.0, Smart Manufacturing, Logistics 4.0, Warehousing 4.0, Internet of Things, Lean Production, Load Balancing, Sorting Conveyor Line, Pharmaceutical Production.

УДК 621.3.095.2 **doi: 10.26906/SUNZ.2024.1.031**

Д. О. Нікітін

Харківський національний університет радіоелектроніки, Харків, Україна

РОЗРОБКА МОДЕЛІ КЕРУВАННЯМ ТЕМПЕРАТУРИ ФОТОПОЛІМЕРНОЇ СМОЛИ НА БАЗІ LCD-ТЕХНОЛОГІЇ 3D-ДРУКУ

Анотація. У статті розглянуто вплив температури фотополімерної смоли для адетивного 3D- друку на якість готових деталей. Розроблена система контролю температури фотполімерної смоли для зменишиння температурного коєфіцієнту об'ємного розширення речовини під час виготовлення обемної моделі. Проаналізовано особливості фотполімерного 3D-друку за технологією LCD. Розглянуто процес виникнення перегріву фотополімера в процесі друку, та розглянуті чиники які впливають на нагрів фотополімерної смоли. За результатами тестів можливо зробити висновки що, контроль температури фотополімерної смоли дозволяє зменшити відхилення геометричних розмірів зменьшилося на 0,013 мм.

Ключові слова: фотополімерний 3D-друк, LCD технологія, фотополімерна смола, теплові тотоки, контроль температури, об'ємне розширення рідини.

Вступ

Розвиток адитивних технологій виробництва все більше стає поширеним в промисловості та повсякденному житі. Одним з най більш універсальним та доступним для споживачів засобом для отримання об'ємних деталей складних форм є 3D-друк [1].

Фотополімерний 3D-друк, являється одним з доступнішим та точним методом протипування деталей. Фотоплімерний друк використовується багатьох сферах, як: ювелірне виробництво (створення майстер-моделей для лиття), стоматологи (створення протезів зубів), створення декоративних моделей для дизайну інтер'єру. Всф фотополімерні технології (SLA, DLP та LCD), працюють на принципі фотополімеризації.

Фотополімеризація – це метод, в якому використовується світло (видиме або ультрафіолетове) для створення хімічної реакції, в результаті якої рідкий матеріал – полімер, стає твердішим у результаті процесу затвердіння

Точність виготовлення деталей за цими технологіями залежить не тільки від технічних характеристик принтера, а і від властивостей фотополімерної смоли. Тому визначення впливу смол на збереження геометричних розмірів, є актуальною задачею при виготовлені моделі [2, 3].

Аналіз мережного тракту

Фотополімерному 3D друкці можна ідентифікувати наступні основні фактори, що впливають на відхилення геометричних розмірів моделі. Для зручності їх можна поділити на дві групи:

– параметри налаштування експонування шарів моделі;

– фізико-хімічні властивості фотополімерної смоли.

Параметри налаштування експозиції шарів моделі, до яких відносяться значення, встановлені в програмі для підготовки моделі до друку (наприклад, NanoDLP або Chitubox). Серед них можна виділити: висота шару, мкм; кількість базових шарів; час експонування базових та основних слоїв, секундах; інтенсивність випромінювання, Лм.

Характеристики моделі та збереження геометричних розмірів під час друку залежать від фізико-хімічних властивостей фотополімерної смоли, що обумовлені її хімічним складом. Серед цих властивостей можна відзначити: коефіцієнт усадки смоли, %; довжина хвилі поляризації, нм; коефіцієнт теплопровідності, Вт/м^2.

Детальне розглядання принципу роботи фотополімерних технологій вказує на певну системність. У всіх трьох цих технологіях використовується випромінювання світла, енергія якого не лише спрямована на полімеризацію необхідних областей фотополімеру, але також на нагрівання самої фотополімерної смоли, зокрема за рахунок температурного коефіцієнта об'ємного розширення матеріалу (ТКОР). Температурний коефіцієнт об'ємного розширення матеріалу α визначає, як змінюється його об'єм при зміні температури і вимірюється в одиницях 1/°C (або 1/K). Він показує, на скільки одиниць зміниться об'єм матеріалу при зміні температури на один градус Цельсія [4].

Математично температурний коефіцієнт об'ємного розширення.

$$\alpha = \frac{\Delta v}{v_0} / \Delta T, \quad (1)$$

де α – температурний коефіцієнт об'ємного розширення, Δv – зміна об'єму матеріалу, v_0 – початковий об'єм матеріалу, ΔT – зміна температури.

Температурний коефіцієнт об'ємного розширення може варіюватися в залежності від конкретного матеріалу. Вплив ТКОР під час формування кожного шару моделі виявляється дуже важливим, оскільки при занадто високій температурі смоли збільшується коефіцієнт розширення матеріалу.

Це може призводити до геометричних відхилень у шарах моделі, а також викликати проблеми, подібні до перезасвідчення шарів, зображених на рис. 1. Отже, можна врахувати, що всі ці параметри, до певної міри, безпосередньо впливають на температуру нагріву смоли під час друку і, отже, на ТКОР. В результаті вивчення температурних впливів у процесі фотополімерного 3D-друку визначається актуальність досліджень.

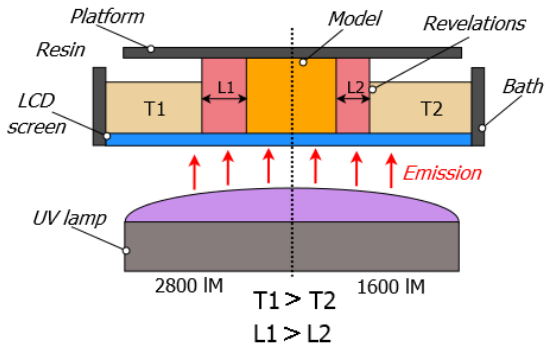


Рис. 1. Принцип впливу ТКОР в процесі експонування моделі

Фізичний опис процес нагріву фотополімерної смоли

З фізичної точки зору можна розглядати процес нагріву фотополімерної смоли та виникнення температурного коефіцієнта об'ємного розширення як проблему теплообміну у тришаровій стінці.

Розглядаючи особливості фотополімерного 3D-друку за технологією LCD, детально розглянемо структуру тришарової стінки, рис. 2. Ця стінка складається з трьох щільно прилягаючих один до одного шарів з такими товщинами: d_1 (товщина LCD екрану), d_2 (товщина плівки) та d_3 (товщина рідкої фотополімерної смоли). Кожен з цих шарів має власну теплопровідність (λ_1, λ_2 і λ_3 відповідно).

Також відомі температури зовнішніх поверхонь t_1 і t_4. Тепловий контакт між шарами є ідеальним, без взаємних зазорів і, відповідно, без повітряних проміжків. Температури в місцях контакту шарів позначаємо як t_2 і t_3.

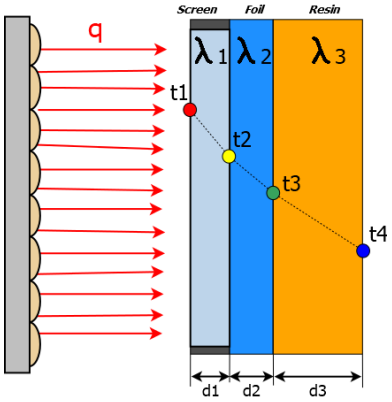


Рис. 2. Тепло-обмін в процесі фотополімерного LCD 3D-друку

Оскільки температури зовнішніх поверхонь постійні, тепловий потік – сталий, і відповідно, кількість теплоти, що проходить за одиницю часу, незмінна [5]. За стаціонарного режиму питомий тепловий потік q постійний і для всіх шарів однаковий. Тому можна записати для кожного з шарів:

$$q = \frac{\lambda_1}{d_1}(t_1 - t_2); \qquad q = \frac{\lambda_2}{d_2}(t_2 - t_3); \quad q = \frac{\lambda_3}{d_3}(t_3 - t_4). \tag{2}$$

З наведених виразів легко визначити значення локальних різниць температур на межах кожного шару:

$$t_1 - t_2 = q\frac{d_1}{\lambda_1}; \quad t_2 - t_3 = q\frac{d_2}{\lambda_2}; \quad t_3 - t_4 = q\frac{d_3}{\lambda_3}. \tag{3}$$

Складаючи по черзі ліві та праві частини рівнянь отримаємо:

$$t_1 - t_4 = q\left(\frac{d_1}{\lambda_1} + \frac{d_2}{\lambda_2} + \frac{d_3}{\lambda_3}\right). \tag{4}$$

Звідки отримуємо рівня теплового потоку:

$$q = \frac{t_1 - t_4}{\frac{d_1}{\lambda_1} + \frac{d_2}{\lambda_2} + \frac{d_3}{\lambda_3}} = \frac{\lambda_{ек}}{d}(t_1 - t_4). \tag{5}$$

Температури на стику шарів t_2 та t_3 можна визначити із системи рівнянь:

$$t_2 = t_1 - q\frac{d_1}{\lambda_1}; \quad t_3 = t_4 + q\frac{d_3}{\lambda_3}. \tag{6}$$

Іноді багатошарову стінку розраховують як одношарову товщиною $d_{заг}$, де $d_{заг}$ – як сума всіх товщин шару. При цьому в розрахунок вводиться еквівалентний коефіцієнт теплопровідності $\lambda_{ек}$, який визначається:

$$q = \frac{t_1 - t_4}{\frac{d_1}{\lambda_1} + \frac{d_2}{\lambda_2} + \frac{d_3}{\lambda_3}} = \frac{\lambda_{ек}}{d}(t_1 - t_4) \tag{7}$$

Звідки отримуємо рівня еквівалентний коефіцієнт теплопровідності:

$$\lambda_{ек} = \frac{d}{\frac{d_1}{\lambda_1} + \frac{d_2}{\lambda_2} + \frac{d_3}{\lambda_3}}. \tag{8}$$

Під час виведення розрахункової формули для багатошарової стінки (7 та 8) припускали, що шари щільно прилягають один до одного і завдяки ідеальному тепловому контакту поверхні, які дотикаються, мають однакову температуру. При шорсткій поверхні між шарами виникають повітряні зазори. А оскільки теплопровідність повітря в нормальних умовах λповітря дорівнює 0,025 Вт/(м*˚С), то наявність навіть дуже тонких повітряних прошарків різко погіршує теплопровідність конструкції.

Розробка моделі керуванням температури фотополімерної смоли

Першим етапом для моделювання впливу теплових процесів на ТКОР фотополімерної смоли, необхідно визначити сам об’єкт керування та що саме можна вважати керуючим впливав для даної моделі.

LCD технологія. Принтер із засвіченням фотополімера світлодіодним УФ-матрицею з використанням в якості маски LCD-дисплея, рис. 3.

Рідка смола або інший світлочутливий матеріал розташовується на поверхні платформи. Після цього джерело УФ-впромінювання, використовується для висвітлення шару рідини, де матеріал стає твердим або полімеризується [3].

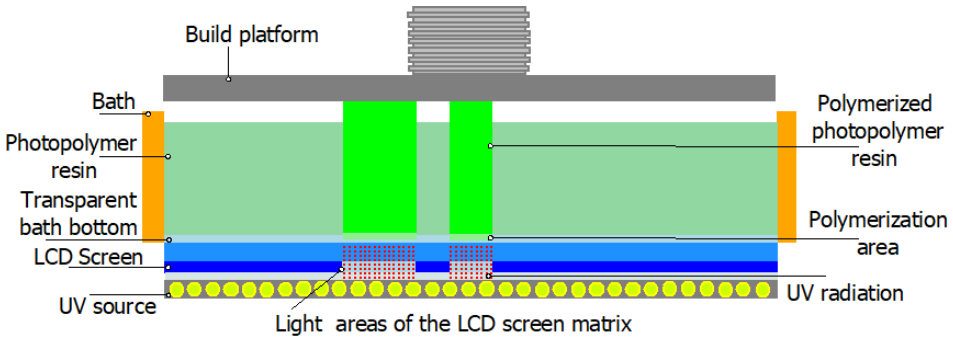


Рис. 3. Схема роботи LCD принтера

УФ матриця задається генератором імпульсів, а довжина та скаженість імпульсів задається за допомогою час засвічення та час циклу. Амплітуда імпульсу будемо вимірювати через силу світла в (кандели: кд×ср×Вт$^{-1}$).

Частина світла, проходячи до робочої зони, перетворюється на тепло в LCD матриці, плівці ванни і фотополімері. Відсоток (коефіцієнт) поглинання залежить від оптичної прозорості та товщини цих шарів. Коефіцієнт визначається експериментально, товщина відома. Світло, що потрапило на стіл або модель, що друкується, фокусується на ньому і бере участь у полімеризації фотополімеру.

Вважаємо, що світло, яке залишилося, повністю витрачається на полімеризацію смоли. Коефіцієнтом відбиття від заготівки та поверхні деталі нехтуємо (поверхня заготівки матова, смола, що полімеризується, має малий коефіцієнт відбиття). Також на процес теплообміну в верстаті та нагрів ті охолодження фотополімерної смоли впливає примусова вентиляція електронних компонентів пристрою, що теж необхідно враховувати. Враховуємо взаємне нагрівання шарів. На рис. 4 наведена схема процесу передачі тепла [5].

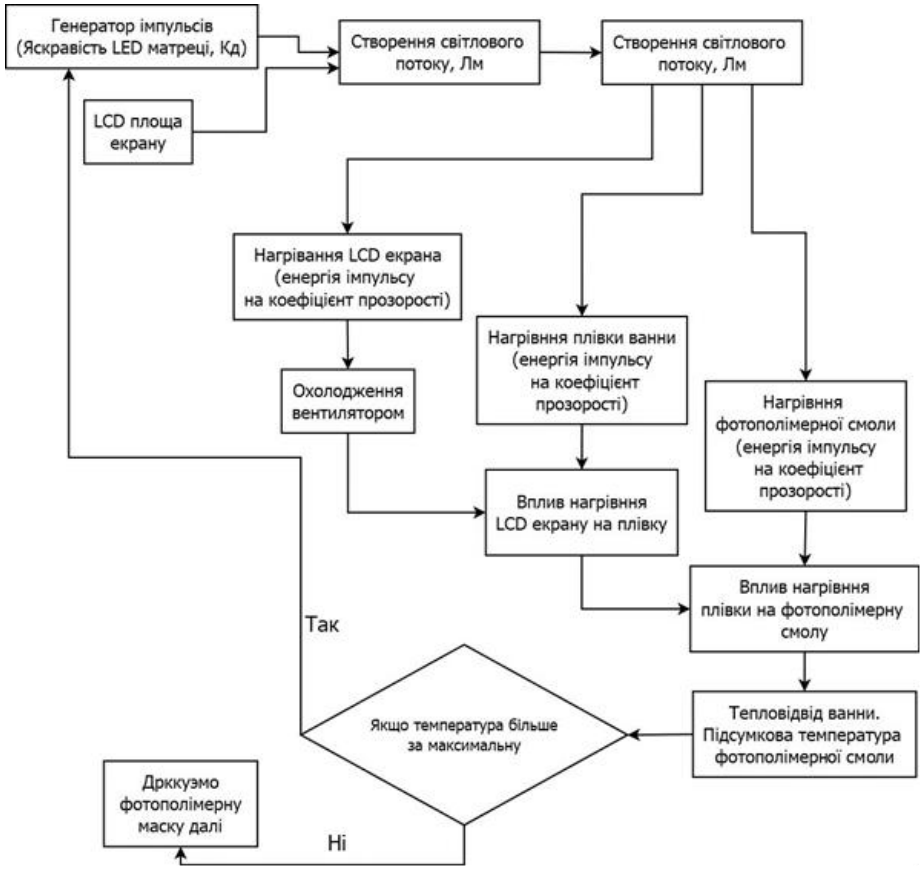


Рис. 4. Алгоритм роботи при експонуванні топології ДП

До параметрів верстату можливо віднести:

– об'єм зони корпусу верстату;

– елементи які нагріваються при експонуванні топології ДП (LCD екран, плівка та фотополімерна смола), (теплоємність та тепло супротив цих елементів).

Дані елементи можна розглянути як стінки. В даній моделі буде використовуватися розрахунок з низькою теплоємністю, оскільки товщина стінок мала. Для створення стін з низькою теплоємністю у моделі було обрано значення повного опору (імпеданс) і ємнісного опору, наведені в табл. 1.

Таблиця 1 – **Розрахунок повного та ємкісного тепло супротиву елементів моделі**

Найменування	Товщина шару, м	Теплоємкість, С, кДж/м²×К	Тепловий супротив R, м²×К/Вт	Питома теплоємність Ср, кДж/кг×К	Питома маса, кг/м²
LCD екран (скло)	0,00135	4,82	0,32	0,84	0,91
Плівка (полімерна)	0,000125	1,8	0,75	0,95	0,9
Фотополімерна смола (епоксидна смола)	0,015	0,17	0,026	1,2	1,16
Лінза УФ-матриці	0,010	4,82	0,32	0,42	0,91
Повітряний прошарок	0,025	0	0,160	0,00	0,00
Кількість діодів в УФ-матриці	24				
Площа дна ванни, мм²	8160				
Об'єм відсіку для електроніки верстату, м³	0,013				

З точки зору автоматизованого контролю процесу, значення ТКОР фотополімерної смоли можливо описати за допомогою рис. 5.

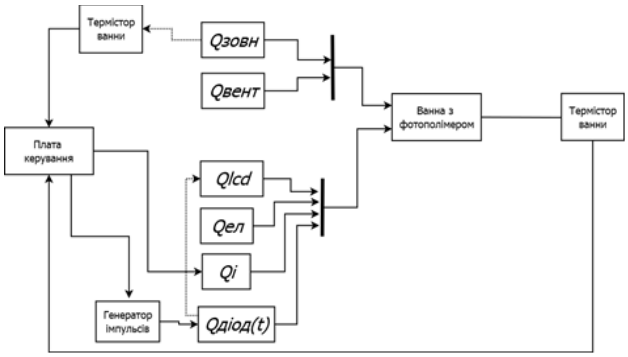


Рис. 5. Принцип роботи керуючого сигналу

Побудова математичної моделі теплових потоків, які впливають на нагрів фотополімерної смоли

В даному випадку на систему нагріву фотополімерної смоли впливає два типи теплових потоків: 1) зовнішній нагрів на макет (температура в приміщенні, природня вентиляція в приміщенні); 2) внутрішній нагрів в середні макету (нагрів від електроніки макету, нагрів від діодів УФ матриці).

Зовнішніх теплових потоків в системі є два:

– $Q_{зовн}$ – кондукційний теплообмін скрізь стінки приладу;

– $Q_{вент}$ – теплообмін від природньої вентиляції.

Математичний опис $Q_{зовн}$, можливо виразити:

$$Q_{\text{зовн}} = \frac{\Delta T}{\Sigma R_{\text{тс}}} = US(T_{\text{в}} - T_{\text{з}}). \quad (9)$$

де R_{TC} – термічний опір, К/Вт; U – загальний термічний опір, Вт/м²×К; S – площина дна ванни для фотополімерної смоли, мм²; $T_З$ – температура зовні верстату, °С; $T_В$ – температура в середні верстату, °С.

Загальний термічний опір знаходимо:

$$U = \frac{1}{\frac{1}{h_i} + \Sigma\frac{L}{\lambda} + \frac{1}{h_e}}. \quad (10)$$

де L – товщина шару, м; λ – коефіцієнт теплопровідності, Вт/м²×К; $h_з$ та $h_в$ – коефіцієнт внутрішнього і зовнішнього конвективного теплообміну відповідно, Вт/м²×К. Математичний опис $Q_{вент}$, можливо виразити як

$$Q_{\text{вент}} = \Phi\rho_{\text{п}}c_{\text{п}}(T_{\text{в}} - T_{\text{з}}). \quad (11)$$

де Φ – потік повітря в наслідок природньої вентиляції, м³/с; $\rho_п$ – щільність повітря (1,2 кг/м³); $c_п$ – питома теплоємкість повітря (1000 кДж/кг).

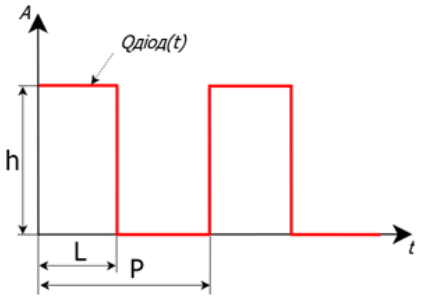
Потік повітря в наслідок природньої вентиляції Φ, знаходимо за допомогою формулою 3.12.

$$\Phi = \frac{nV}{3600}. \quad (12)$$

де V – об'єм повітря в середні верстату, м³.

До внутрішніх чинників теплообміну можна віднести: Q_{LCD} – загальний внутрішній нагрів від УФ-матриці; $Q_{ел}$ – внутрішній нагрів від електронних компонентів верстата; Q_i – потік тепла внаслідок інфільтрації (примусова вентиляція); $Q_{діод}(t)$ – тепло від одного УФ-діода.

За допомогою Q_{LCD}, Q_i та $Q_{діод}(t)$, можливо контролювати ТКОР фотополімерної смоли. Основним елементом нагріву є УФ-матриця, яка генерує УФ-випромінювання. Для зниження можливості перегріву фотополімеру, вмикання діодів на УФ-матриці можливо зробити за допомогою імпульсів тим самим робити інтервали між вмиканням та вимиканням діодів. Це буде дозволяти дати час смолі охолонути та зменшити імовірність перегріву та виникнення великого значення ТКОР. Отже автоматизований контроль буде відбуватися за допомогою керуванням теплового потоку $Q_{діод}(t)$, рис. 6.



A – амплітуда імпульсу; *L* – довжина імпульсу; *P* – період імпульсу.

Рис. 6. Принцип роботи керуючого сигналу

Опис $Q_{діод}(t)$ можна зробити таким чином:

$$Q_{\text{діод}}(t) = q_{\text{випром}}S_{\text{лінза}}SGF. \quad (13)$$

де $q_{випром}$ – випромінювання діода, Вт/м²; $S_{лінзи}$ – площа лінзи для фокусування УФ-випромінювання, мм²; SGF – коефіцієнт посилення.

Випромінювання діода $q_{випром}$ знаходимо:

$$q_{\text{випром}} = \tau(E_{\text{пр}} + E_{\text{роз}})SGF. \quad (14)$$

де τ – передана енергія; $E_{пр}$ – пряма енергія випромінювання від діоду, Вт/м^2; $E_{роз}$ – пряма енергія випромінювання від діоду, Вт/м^2.

Загальний внутрішній нагрів від УФ-матриці Q_{LCD}, можна виразити:

$$Q_{LCD} = H_д n_д. \quad (15)$$

де $H_Д$ – норма виробітки тепла діода, Вт; $n_Д$ – кількість діодів на УФ-матриці.

Внутрішній нагрів від електронних компонентів верстата $Q_{ел}$, будемо вважати як сталу величенну. Потік тепла внаслідок інфільтрації (примусова вентиляція) Q_i, знаходимо:

$$Q_{\text{і}} = 0{,}013 V_r.(T_{\text{в}} - T_{\text{з}}). \quad (16)$$

де 0,013 – це об'єм нагрітого повітря в середні верстату, Дж/м^3; Vr – рівень вентиляції, м^3/с.

Диференціальне рівнянь, яке буде включати всі теплові потоки та пов'язувати їх вплив з температурою фотополімеру буде мати наступний вигляд:

$$m_{\text{см}} c_{\text{см}} \frac{dT}{dt} = Q_{\text{діода}}(t) + Q_{LCD} + Q_{\text{ел}} + Q_{\text{зовн}} - Q_{\text{вент}} - Q_{\text{і}}. \quad (17)$$

де $m_{см}$ – фотополімерної смоли в ванній, кг; $c_{см}$ – питома теплоємкість фотополімерної смоли.

Динамічна модель можливо виразити за допомогою електор-теплової аналогії. В схемі є 5 резисторів (середовища через які відбуваються втрати теплових потоків) та чотири конденсатори (середовища які поступово нагріваються в процесі експонування зображення топології) [6, 7], рис. 7.

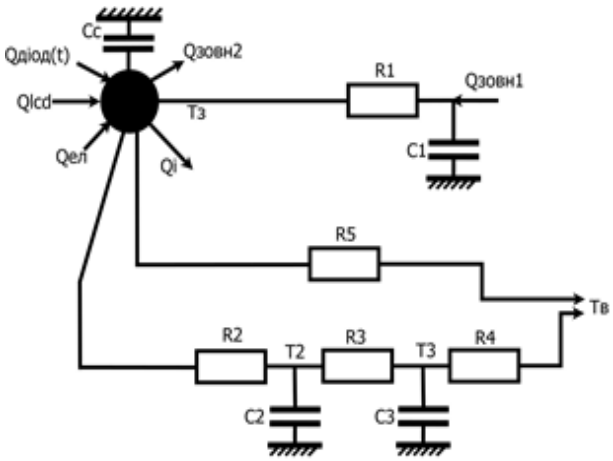


Рис. 7. Електоро-теплова схема теплообміну

За даною схемою виходить що: $Q_{зовн1}$ – приплив теплоти безпосередньо від переміщення, Вт; $Q_{зовн2}$ – приплив теплоти від корпусу верстата, Вт; $Q_{вент}$ – приплив теплоти від природньої вентиляції, Вт; Q_{LCD} – приплив теплоти від загального внутрішнього нагріву від УФ-матриці, Вт; $Q_{ел}$ – приплив теплоти від електронних компонентів верстата, Вт; Q_i – приплив теплоти внаслідок інфільтрації повітря, Вт; $Q_{діод(t)}$ – приплив теплоти від одного УФ-діода, Вт; T_3 – температура зовні верстату, °С; T_B – температура в середні верстату, °С; T_2 – температура між УФ-матрицею та повітрям, °С; T_3 – температура між повітрям та LCD екраном, °С; C_c – теплоємність фотополімерної смоли в ванній, Дж/К; C_1 – теплоємність матеріалів корпусу верстату, Дж/К; C_2 – теплоємність фіксуючої лінзи, Дж/К; C_3 – LCD-екрану, Дж/К; R_1 – опір матеріалів корпусу, К/Вт; R_2 – опір матеріалу лінзи, К/Вт; R_3 – опір повітря між лінзою та LCD-екраном, К/Вт; R_4 – опір LCD-екрана, К/Вт; R_5 – сумарний термічний опір.

Динамічний режим верстата можна змоделювати за допомогою системи звичайних лінійних диференціальних рівнянь. Ці рівняння можна переписати у вигляді матриць, а потім реалізувати динамічну модель у середовищі MatLab/Simulink. Рівняння для температурних вузлів мають такий вигляд [8]:

$$\begin{aligned} C_1 \frac{dT_1}{dt} &= \frac{1}{R_1}(T_з - T_1) + q_{\text{вирпом}} S_{\text{лінза}} SGF\vartheta; \\ C_2 \frac{dT_2}{dt} &= \frac{1}{R_2}(T_з - T_2) - \frac{1}{R_3}(T_2 - T_3); \\ C_3 \frac{dT_3}{dt} &= \frac{1}{R_3}(T_2 - T_3) - \frac{1}{R_4}(T_3 - T_4). \end{aligned} \quad (18)$$

де показник конвекції:

$$\begin{aligned} C_{\text{с}} \frac{dT_c}{dt} &= Q_{\text{діод}}(t) + Q_{LCD} + Q_{\text{ел}} \\ &+ q_{\text{випром}} S_{\text{лінза}}(1 - \vartheta) + mcT_{\text{з}} \\ &- \frac{1}{R_1}(T_{\text{з}} - T_1) - \frac{1}{R_2}(T_{\text{з}} - T_2) - \frac{1}{R_5}(T_{\text{з}} - T_{\text{в}}). \end{aligned} \quad (19)$$

Якщо $\vartheta = 0$, то енергія випромінювання повністю передається через пікселі LCD-екрана. Якщо $\vartheta = 1$, то вся енергія випромінювання безпосередньо нагріває фотополімерну смолу.

Після перетворення рівняння мають такий вигляд:

$$\begin{aligned} \dot{T}_1 &= \left(\frac{1}{R_1 C_1}\right) T_i + \left(-\frac{1}{R_1 C_1}\right) T_1 + \left(\frac{S_{\text{лінзи}} SGF\vartheta}{C_1}\right) q_{\text{випром}}; \\ \dot{T}_2 &= \left(\frac{1}{R_2 C_2}\right) T_i + \left(-\frac{1}{R_3 C_3} - \frac{1}{R_3 C_2}\right) T_2 + \left(\frac{1}{R_3 C_2}\right) T_3; \\ \dot{T}\dot{T}_3 &= \left(-\frac{1}{R_3 C_3}\right) T_2 + \left(-\frac{1}{R_3 C_3} - \frac{1}{R_4 C_3}\right) T_3 + \left(\frac{1}{R_4 C_3}\right) T_{\text{в}}; \\ \dot{T}_{\text{в}} &= \left(\frac{1}{C_{\text{с}}}\right) Q_{\text{діод}}(t) + \left(\frac{1}{C_{\text{с}}}\right) Q_{LCD} + \left(\frac{1}{C_{\text{с}}}\right) Q_{\text{ел}} \\ &+ \left(\frac{1}{C_{\text{с}}}\right) [q_{\text{изл}} S_{\text{стекл}} SGF(1 - \vartheta)] + \\ &\left(\frac{\dot{m}c}{C_{\text{с}}} + \frac{1}{R_5 C}\right) T + \left(\frac{1}{R_1 C_{\text{с}}}\right) T_1 + \left(\frac{1}{R_2 C_{\text{с}}}\right) T_2 + \\ &\left(\frac{\dot{m}c}{C_{\text{с}}} - \frac{1}{R_2 C_{\text{с}}} - \frac{1}{R_5 C_{\text{с}}} - \frac{1}{R_1 C_{\text{с}}}\right) T_{\text{в}}. \end{aligned} \quad (20)$$

Рівняння 19 та 20 можна подати в матричному вигляді:

$$\begin{aligned} \dot{x} &= Ax + Bu; \\ y &= Cx + Du. \end{aligned} \quad (21)$$

З такими векторами отримуємо наступну матрицю рівнянь:

$$x = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_i \end{bmatrix}; \quad \dot{x} = \begin{bmatrix} \dot{T}_1 \\ \dot{T}_2 \\ \dot{T}_3 \\ \dot{T}_i \end{bmatrix}; \quad u = \begin{bmatrix} T_{\text{в}} \\ Q_{\text{вент}} \\ Q_{\text{зовн}} \\ q_{\text{випром}} \end{bmatrix}; \quad y = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_i \end{bmatrix};$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}; \qquad D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix};$$

$$A = \begin{bmatrix} \frac{-1}{R_1 C_1} & 0 & 0 & \frac{1}{R_1 C_1} \\ 0 & \frac{-1}{R_1 C_1} - \frac{1}{R_3 C_2} & \frac{1}{R_3 C_2} & \frac{1}{R_2 C_2} \\ 0 & \frac{1}{R_3 C_3} & \frac{-1}{R_3 C_3} - \frac{1}{R_4 C_3} & 0 \\ \frac{1}{R_1 C_c} & \frac{1}{R_1 C_1} & 0 & -\frac{\dot{m}c}{C_c} - \frac{1}{R_2 C_c} \\ & & & -\frac{1}{R_5 C_c} - \frac{1}{R_1 C_c} \end{bmatrix};$$

$$B = \begin{bmatrix} 0 & 0 & 0 & \frac{S_{\text{лінза}} SGF\vartheta}{C_1} \\ 0 & 0 & 0 & 0 \\ \frac{1}{R_4 C_3} & 0 & 0 & 0 \\ \frac{\dot{m}c}{C_c} + \frac{1}{R_5 C_c} & \frac{1}{C_c} & \frac{1}{C_c} & \frac{S_{\text{лінза}} SGF\vartheta}{C_c} \end{bmatrix}. \qquad (22)$$

Дослідження теплових режимів фотополімерної смоли виконується шляхом розробки динамічної моделі в середовищі MatLab/Simulink. Реалізація моделі в середовищі Simulink наведена, на рис. 8.

Розташування елементів в конструкції LCD принтера впливає на температуру повітря всередині верстата. Для контроля температури фотополімерної смоли під час експонування зображення. В верстат буде встановлено два термістора: термістор контролю зовнішньої температури; термістор контролю температури фотополімерної смоли. Для контроля сигналів імпульсів УФ-лампи верстату буде використовуватися за допомогою PID-регулювання (пропорційно-інтегрально-диференціювальний регулятор) [5, 6]. Отже, за допомогою PID-регулятору буде відбуватися порівння температури фотополімерної смоли з значенням температури вказаної на контролері, кожні 1000 мс. Зазначення на контролері, вказується власноруч, або воно буде обиратися автоматично, виходячи з коливань зовнішньої температури навколо макету.

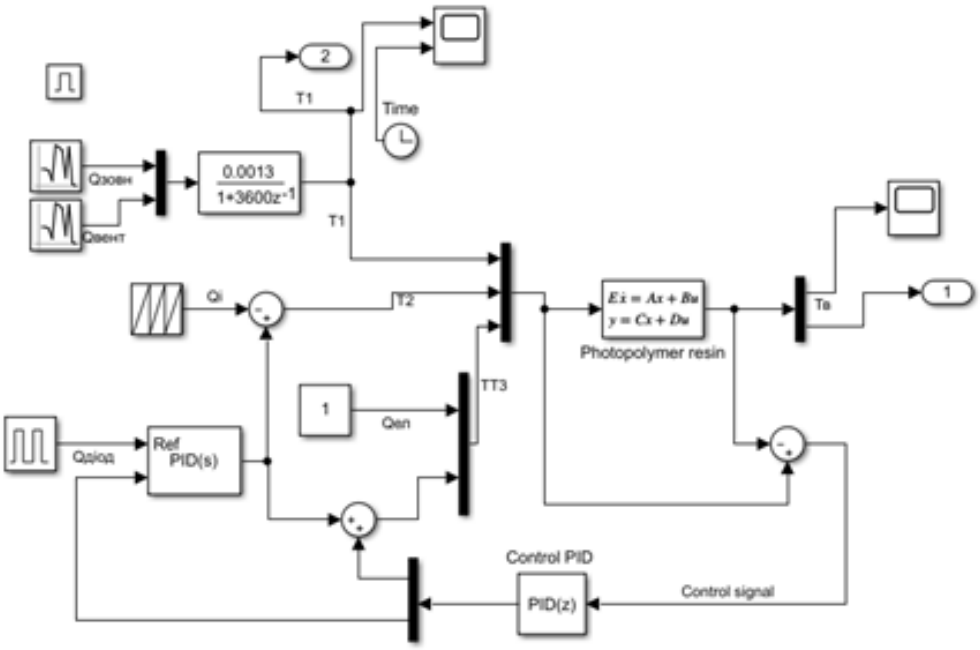
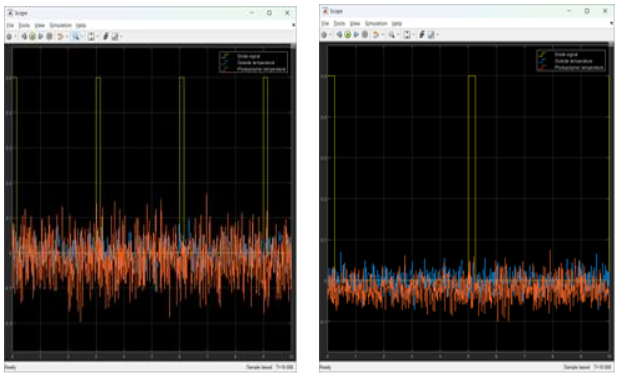


Рис. 8. Схема контролю температури фотополімерної смоли

Якщо значення температури фотополімерної смоли під час експонування зображення більше ніж значення на контролері, то PID-регулятор передає сигнал контролери. Після цього контролер вносить корективи в G-code, та робить більший інтервал між включеннями УФ-матриці. Тим самим збільшується час на охолодження фотополімерної смоли. Після чого знову відбувається порівння температури.

Якщо коливання тепла в фотополімерної смоли не перевищують температури вказану температури то продовжується друг, якщо більше то знову збільшується час між вмиканням та вимиканням УФ-матриці. На рис. 9 наведені результати роботи схеми контролю температури фотополімерної смоли.

Для перевірки роботи розробленої схеми контроль температури фотополімерної смоли, створено дві тестові моделі розмірами 20×20×20 мм.

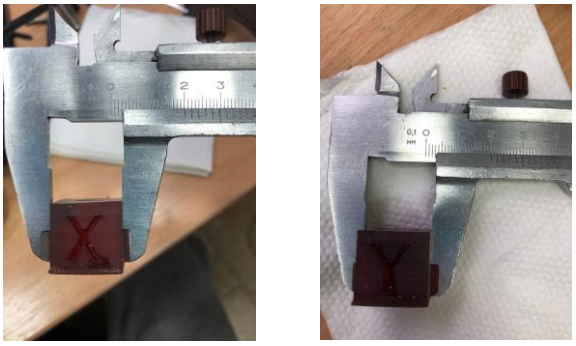


а б

Рис. 9. Результати роботи схеми контролю температури фотополімерної смоли, температура при роботі верстата:
а – без контроля з перюрдом експонування 3 с;
б – з контролем з перюрдом експонування 6 с

Параметри друку наступні:
– тривалість засвічення смоли 8 секунд;
– інтенсивність випромінювання 1600 Лм;
– довжина хвилі випромінювання 435 нм;
– товщина базового шару 35 мкм.

Одна модель буде зроблена без регуляції температури інша з контролем температури фотополімерної смоли. Результати друку наведені на рис. 10.



а б

Рис. 10. Тестові зразки: а – з контролем температури фотополімерної смоли; б – без контролю температури

За результатами тестів можливо зробити висновки що, при контролю температури фотополімерної смоли, відхилення геометричних розмірів зменьшилося на 0,013 мм.

Висновки

У статті розглянуто вплив температури фотополімерної смоли на відхилення геометричних розмірів під час LCD 3D-друку. Були розглянуті наступні питання:

1. Фізичний опис процес нагріву фотополімерної смоли в процесі експонування моделі. Були розглянуті особливості фотополімерного експонування за LCD технологією, а саме процес передачі теплового потоку від УФ випромінювання діодної матриці до фотополімерної смоли. Тим саме пояснюється описується виникнення теплового коефіцієнту об’ємного розширення фотополімерної смоли підчас друку моделі;

2. Розробка моделі впливу температури в Matlab Simuling. Була розроблена схема керування тепловим потоком при фотополімерному експонуванні на основі задачі теплообміну між трьома стінками з різними коефіцієнтами теплопровідності. Описані формули та побудована електрична схема нагріву за яким відбуваються процес нагріву фотополімерної смоли. Та проведення проведені симуляції в програмному середовищі Matlab Simuling.

За результатами тестів можливо зробити висновки що, контроль температури фотополімерної смоли дозволяє зменшити відхилення геометричних розмірів зменьшилося на 0,013 мм.

Список літератури

1. Fiedor, P.; Pilch, M.; Szymaszek, P.; Chachaj-Brekiesz, A.; Galek, M.; Ortyl, J. Photochemical Study of a New Bimolecular Photoinitiating System for Vat Photopolymerization 3D Printing Techniques under Visible Light. Catalysts 2020, 10, 284.
2. Nevlyudov I., Razumov-Fryziuk I., Nikitin D., Blyzniuk D., Strelets R. Technology for creating the topology of printed circuit boards using polymer 3D masks // № 1 (15) (2021): Сучасний стан наукових досліджень і технології в промисловості. ст 120-131. DOI: https://doi.org/10.30837/ITSSI.2021.15.120.
3. Нікітін Д.О., Стрілець Р.Є., Близнюк Д.С. Порівняльний аналіз технологій 3D прототипування SLA, DLP та LCD. Розробка автоматизованої станції для 3D друку // Матеріали VII Міжнародної науково-технічної Internet-конференції «Сучасні методи, інформаційне, програмне та технічне забезпечення систем керування організаційно-технічними та технологічними комплексами», 26 листопада 2020. [Електронний ресурс] – К: НУХТ, 2020. 55 – 56 с.
4. Теорія автоматичного управління: Навчальний посібник [Електр. ресурс]: уклад.: О. Й. Штіфзон, П. В. Новіков, В.П. Бунь. – Електронні текстові дані (1 файл: 2,2 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2020. – 144 с.
5. Vsevolod M, Kuntsevich, Vyacheslav F. Gubarev, Yuriy P. Kondratenko, Dmitriy V. Lebedev, Vitaliy P. Lysenko. Control Systems: Theory and Applications. CRC Press 2022, p. 327. ISBN 978-87-7022-024-8
6. Конспект лекцій з дисципліни «Теорія автоматичного управління» для здобувачів освітнього ступеня «бакалавр» зі спеціальності 151 Автоматизація та комп`ютерно-інтегровані технології денної форми навчання [Електронний ресурс] / [Упорядник Я. В. Корпань]; Черкас. держ. технол. ун-т. – Черкаси: ЧДТУ, 2019. – 124 с.
7. Методи сучасної теорії управління: підручник / А.П. Ладанюк, Н.М. Луцька, В.Д. Кишенько, Л.О. Власенко, В.В. Іващук – Київ : Видавництво Ліра-К, 2018. – 368 с. ISBN 978-617-7605-36-1
8. Brian H. Hahn, Daniel T. Valentine. “Essential MATLAB for Engineers and Scientists”, Academic Press. ISBN: 978-0-08-100877-5, https://doi.org/10.1016/C2015-0-02182-7. (2017).

Received (Надійшла) 23.11.2023
Accepted for publication (Прийнята до друку) 17.01.2024

Development of a model for controlling the temperature of photopolymer resin based on LCD 3D printing technology

Dmytro Nikitin

Abstract. The article considers the influence of the temperature of photopolymer resin for adhesive 3D printing on the quality of finished parts. A system for controlling the temperature of photopolymer resin has been developed to change the temperature coefficient of volume expansion of a substance during the manufacture of a three-dimensional model. The features of photopolymer 3D printing using LCD technology are analysed. The process of photopolymer overheating in the printing process is considered, and the factors that affect the heating of photopolymer resin are discussed. According to the results of the tests, it is possible to draw conclusions that the control of the temperature of the photopolymer resin allows to reduce the deviation of the geometric dimensions by 0.013 mm.

Keywords: photopolymer 3D printing, LCD technology, photopolymer resin, thermal currents, temperature control, liquid volume expansion.

Інформаційні технології

УДК 007.629.735 **doi: 10.26906/SUNZ.2024.1.038**

А. Н. Аль-Амморі, М. М. Дехтяр, Р. М. Іщенко, А. Є. Клочан

Національний транспортний університет, Київ, Україна

МЕТОДИ ТА ЗАСОБИ ЗАХИСТУ ІНФОРМАЦІЇ

Анотація. Розглядаються загальні питання організації методів і засобів захисту інформації. Розглянуто різні визначення загальнонаукового поняття "інформація", з точки зору різних вчених, дослідників, і залежно від галузі людської діяльності. Розглянуто види представлення інформації та її окремі властивості, стосовно комп'ютерного опрацювання даних. Розглянуто фундаментальні поняття та визначення з області інформаційної безпеки систем. Наведено історичні етапи розвитку засобів захисту інформації, дано класифікацію методів захисту інформації, досліджено основні напрямки їх використання. Розглянуто класифікацію комп'ютерних вірусів за основними їхніми ознаками, а також завдання, які розв'язують антивірусні засоби. Окремо розглянуто криптографічні методи захисту інформації та загальну технологію шифрування.

Ключові слова: інформація, інформаційна безпека, конфіденційність, цілісність, доступність.

Вступ

Протягом усього свого існування людство отримувало нові знання про навколишній світ. Саме нові дані, отримані в процесі пізнання або навчання, називається інформацією. Інформація (лат. informatio - роз'яснення, виклад), першопочатково - відомості, які люди передають усно, письмово або в інший спосіб за допомогою умовних сигналів, технічних засобів тощо. Із середини 20-го століття інформація є загальнонауковим поняттям, що включає: відомості, що передаються між людьми, людиною й автоматом, автоматом і автоматом; сигнали в тваринному і рослинному світі; ознаки, що передаються від клітини до клітини, від організму до організму тощо.

Сучасне наукове уявлення про інформацію дуже точно сформулював Норберт Вінер, "батько" кібернетики: Інформація - це позначення змісту, отриманого із зовнішнього світу в процесі нашого пристосування до нього та пристосування до нього наших почуттів. Відоме також інше його визначення цього поняття: "Інформація є інформація, а не матерія і не енергія". Тим самим Н. Вінер відмовився від формулювання поняття інформації, вважаючи, що воно споріднене з такими категоріями, як рух, життя, свідомість. Добре відоме визначення академіка В.М. Глушкова: "Інформація в найзагальнішому її розумінні являє собою міру неоднорідності розподілу матерії та енергії в просторі та в часі, міру змін, якими супроводжуються всі процеси, що відбуваються у світі. ... Інформацію несуть у собі не тільки поцятковані буквами аркуші книги або людська мова, а й сонячне світло, складки гірського хребта, шум водоспаду, шелест листя". Уявлення про інформацію, що ґрунтуються на статистичній теорії передавання сигналів К. Шеннона, привели до такого визначення у Webster's New World Dictionary of Computer Terms: "Інформація - це дані, які обробляються комп'ютером і можуть бути виведені у формі, зручній для користувача". Визначень інформації існує безліч, причому академік М. М. Моїсеєв навіть вважав, що з огляду на широту цього поняття немає і не може бути суворого і досить універсального визначення інформації. Водночас формулювання терміна "інформація", хоча б у загальному вигляді, необхідне для розв'язання як теоретичних, так і практичних завдань сучасної науки і техніки. Багато в чому визначення інформації залежить від галузі людської діяльності:

- у *побутовому* сенсі під інформацією розуміють будь-які дані або відомості, які когось цікавлять. Наприклад, повідомлення про якісь події, про чиюсь діяльність тощо;

- у *техніці* під інформацією розуміють повідомлення, що передаються у формі знаків або сигналів (у цьому разі є джерело повідомлень, одержувач (приймач) повідомлень, канал зв'язку);

- у кібернетиці під інформацією розуміють ту частину знань, яку використовують для орієнтування, активної дії, управління, тобто з метою збереження, вдосконалення, розвитку системи.

Стосовно *комп'ютерного опрацювання* даних, під інформацією розуміють деяку послідовність символічних позначень (літер, цифр, звуків, графіків, малюнків тощо), яка має смислове навантаження та подана у зрозумілому комп'ютеру вигляді. Фізично інформація в комп'ютері записується і передається у вигляді електричних сигналів. Найзагальніше розуміння терміна "інформація" полягає в тому, що інформація - це відображення розмаїття в існуючому світі. Важливо пам'ятати під час вивчення цього терміна, що жодне з наведених трактувань не може вважатися визначенням. Інформація може існувати у вигляді: тексту, малюнків, фотографій, креслень; світлових або звукових сигналів; радіохвиль; електричних і нервових імпульсів; магнітних записів; жестів і міміки; запахів і смакових відчуттів; хромосом, за посередництвом яких передаються у спадок ознаки і властивості організмів, тощо. Людина сприймає за допомогою органів чуття таку інформацію:

- *візуальну* (сприйняття зорових образів, розрізнення кольорів тощо) – за допомогою зору (90%);

- *звукову* (сприйняття музики, мови, сигналів, шуму тощо) - за допомогою слуху;

- *нюхову* (сприйняття запахів) – за допомогою нюху;

- *смакову* (сприйняття за допомогою смакових рецепторів язика) – за допомогою смаку;

- *тактильну* (за допомогою шкірного покриву сприйняття інформації про температуру, якість предметів тощо) – за допомогою дотику.

Властивості інформації:

- *релевантність* - здатність інформації відповідати потребам (запитам) споживача;

- *повнота* - властивість інформації вичерпно (для даного споживача) характеризувати відображуваний об'єкт або процес;

- *своєчасність* - здатність інформації відповідати потребам у потрібний момент часу;

- *достовірність* - властивість інформації не мати прихованих помилок. Достовірна інформація з часом може стати недостовірною, якщо застаріє і перестане відображати справжній стан справ;

- *доступність* - можливість отримання інформації даним споживачем;

- *захищеність* - властивість, що характеризує неможливість несанкціонованого використання або зміни інформації;

- *ергономічність* - властивість, що характеризує зручність форми або обсягу інформації з точки зору даного споживача.

Основна частина

Всупереч поширеній думці звичайні факти самі по собі ні про що не говорять - вони набувають значення порівняно з іншими фактами. Якщо якісь відомості не несуть для нас смислового навантаження або ж вони не нові для нас, то такий факт залишиться для нас лише фактом. Щодня ми дізнаємося масу нової інформації, потрібної і не дуже. Більшу її частину ми, природно, дізнаємося з Інтернету. Дані та відомості про якісь об'єкти, що перебувають у вільному доступі, належать усім без винятку. Крім того, в Інтернеті люди часто передають один одному величезну кількість особистої інформації. Тому наше суспільство часто називають "інформаційним", адже людство в буквальному сенсі стало залежати від тієї інформації, якою воно володіє.

У сучасному світі інформація являє собою певну для людини цінність. Як і будь-яку іншу цінність, інформацію варто захищати від її неправомірного спотворення або несанкціонованого доступу до неї. Багато користувачів залишилися б незадоволеними, якби інформація особистого характеру, якою вони обмінюються в різних соціальних мережах, перебувала в загальному користуванні. Тому, захист даних від несанкціонованого доступу є одним із пріоритетних завдань під час проектування будь-якої інформаційної системи. Але як правильно захистити інформацію? І чи існує абсолютний захист інформації? Саме ці питання будуть розглянуті в роботі.

Черговий етап технологічної революції, що відбувається нині у світі, спричиняє серйозні зміни в економіці, соціальній структурі суспільства. Масове застосування нових технологічних засобів, на основі яких здійснюється інформатизація, стирає геополітичні кордони, змінює спосіб життя мільйонів людей. Водночас інформаційна сфера стає не тільки однією з найважливіших сфер міжнародного співробітництва, а й об'єктом суперництва.

Нині більшість керівників підприємств і організацій вживають заходів щодо "захисту й оборони" важливої для них інформації. Однак практика показує, що ці дії не завжди мають системний характер. Здебільшого вони спрямовані на ліквідацію тільки окремих загроз, залишаючи проломи в обороні.

На жаль, в Україні до теперішнього часу відсутня єдина система безпеки підприємництва. Тому керівництву будь-якої організації доводиться самостійно вирішувати складне завдання забезпечення своєї економічної та інформаційної безпеки з оптимальними фінансовими витратами, але на необхідному рівні захищеності. Кожне підприємство й організація змушені постійно вести конкурентну боротьбу за своє існування, за прибуткове ведення справ, за своє добре ім'я в умовах становлення ринкової економіки. Успіх виробничої та підприємницької діяльності значною мірою залежить від уміння розпоряджатися таким найцінніш товаром, як інформація. Тому в умовах посилення конкуренції успіх підприємництва, гарантія отримання прибутку дедалі більшою мірою залежать від збереження в таємниці секретів виробництва, що спираються на певний інтелектуальний потенціал і конкретну технологію. Саме поняття "безпека" набуває розширеного змісту, воно охоплює питання інформаційно-комерційної, юридичної та фізичної безпеки, розв'язання яких потребує особливої уваги у зв'язку зі зростаючою роллю інформації в житті суспільства. Розглянемо фундаментальні поняття та визначення з галузі інформаційної безпеки та надійності систем [1– 3]:

Інформація - відомості (дані) про внутрішній і навколишній світ, події, процеси, явища тощо, які сприймаються і передаються людьми або технічними пристроями.

Інформаційна (інформаційно-обчислювальна) система - організаційно впорядкована сукупність документів, технічних засобів та інформаційних технологій, що реалізує інформаційні (інформаційно-обчислювальні) процеси.

Інформаційні процеси - процеси збирання, накопичення, зберігання, опрацювання (перероблення), передавання та використання інформації.

Інформаційні ресурси - окремі документи або масиви документів в інформаційних системах.

Доступ - спеціальний тип взаємодії між об'єктом і суб'єктом, у результаті якого створюється потік інформації від одного до іншого.

Несанкціонований доступ (НСД) - доступ до інформації, пристроїв її зберігання та оброблення, а також до каналів передавання, який реалізують без відома (санкції) власника, порушуючи тим самим встановлені правила доступу.

Об'єкт - пасивний компонент системи, що зберігає, переробляє, передає або приймає інформацію; приклади об'єктів: сторінки, файли, папки, директорії, комп'ютерні програми, пристрої (монітори, диски, принтери тощо).

Суб'єкт - активний компонент системи, який може ініціювати потік інформації; приклади суб'єктів: користувач, процес або пристрій.

Безпека ІВС - властивість системи, що виражається у здатності системи протидіяти спробам неса-

нкціонованого доступу або заподіяння шкоди власникам і користувачам системи за різних навмисних і ненавмисних впливів на неї.

Захист інформації - організаційні, правові, програмно-технічні та інші заходи щодо запобігання загрозам інформаційній безпеці та усунення їх наслідків.

Атака - спроба несанкціонованого подолання захисту системи.

Інформаційна безпека (ІБ) систем - властивість інформаційної системи або реалізованого в ній процесу, що характеризує здатність забезпечити необхідний рівень свого захисту.

Інше визначення:

Інформаційна безпека - усі аспекти, пов'язані з визначенням, досягненням і підтриманням конфіденційності, цілісності, доступності інформації або засобів її обробки:

конфіденційність (confidentiality) - стан інформації, за якого доступ до неї здійснюють тільки суб'єкти, що мають на неї право;

цілісність (integrity) - уникнення несанкціонованої модифікації інформації;

доступність (availability) - уникнення тимчасового або постійного приховування інформації від користувачів, які отримали права доступу.

Цінність інформації визначається ступенем її корисності для власника.

Ідентифікація - процес розпізнавання певних компонентів системи (об'єктів або суб'єктів) за допомогою унікальних ідентифікаторів.

Автентифікація - перевірка ідентифікації користувача або іншого компонента ІС для ухвалення рішення про дозвіл доступу до ресурсів системи.

Надійність системи - характеристика здатності програмного, апаратного, апаратно-програмного засобу виконати за певних умов необхідні функції протягом певного періоду часу за певних умов.

Достовірність роботи системи (пристрою) - властивість, що характеризує істинність кінцевого (вихідного) результату роботи (виконання програми), яка визначається здатністю засобів контролю фіксувати правильність або помилковість роботи.

Помилка пристрою - неправильне значення сигналу (біта - у цифровому пристрої) на зовнішніх виходах пристрою або окремого його вузла, спричинене технічною несправністю, або перешкодами, що впливають на нього (навмисними чи ненавмисними), або іншим способом.

Помилка програми - проявляється в невідповідному реальному (необхідному) проміжному або кінцевому значенню (результату) внаслідок неправильно запрограмованого алгоритму або програми.

Достовірність інформації визначається достатньою для володільця точністю відображати об'єкти і процеси навколишнього світу в певних часових і просторових рамках. Інформація, що спотворено представляє дійсність, може завдати власнику значної матеріальної та моральної шкоди. Якщо інформація спотворена навмисно, то її називають дезінформацією. Своєчасність інформації, тобто відповідність цінності та достовірності певному часовому періоду. Ця властивість визначається виразом

$$C(t) = C_0 e^{-2,3t/\tau},$$

де C_0 – цінність інформації в момент її виникнення; t – час від моменту виникнення інформації до моменту визначення її вартості; τ – час від моменту виникнення інформації до моменту її застарівання. Історичні етапи розвитку засобів захисту інформації представлено табл. 1 [4–6].

Таблиця 1 – **Історичні етапи розвитку засобів захисту інформації**

Етапи розвитку засобів ІБ	Коротка характеристика етапу
I етап - приблизно до 1915/16 року	характеризується використанням засобів інформаційних комунікацій, що природно виникали. Основне завдання інформаційної безпеки - захист відомостей про події, факти, майно тощо.
II етап – починаючи з 1916 року	пов'язаний із початком використання технічних засобів електро- і радіозв'язку. Характеризується застосуванням завадостійкого кодування повідомлення (сигналу) з подальшим декодуванням прийнятого повідомлення (сигналу).
III етап – починаючи з 1935 року	пов'язаний із появою радіолокаційних і гідроакустичних засобів. Забезпечення інформаційної безпеки ґрунтувалося на поєднанні організаційних і технічних заходів, спрямованих на підвищення захищеності радіолокаційних засобів від впливу на їхні приймальні пристрої активних і пасивних перешкод.
IV етап – починаючи з 1946 року	пов'язаний із винаходом і впровадженням у практичну діяльність електронно-обчислювальних машин (комп'ютерів). Еру появи комп'ютерної техніки пов'язують із розробкою в Пенсільванському університеті (США) ЕОМ EN IAC (Electronic Numerical Integrator And Computer (Calculator)). Завдання інформаційної безпеки розв'язували здебільшого методами та способами обмеження фізичного доступу до обладнання засобів збирання, перероблення та передавання інформації.
V етап – починаючи з 1964/65 років	обумовлений створенням та розвитком локальних інформаційно-комунікаційних мереж. Завдання безпеки вирішувалися в основному методами та способами фізичного захисту коштів шляхом адміністрування та управління доступом до мережних ресурсів.
VI етап - починаючи з 1973 року	пов'язаний із використанням мобільних комунікаційних пристроїв із широким спектром завдань. У цей період створено відомі зараз у всьому світі фірми Microsoft (Білл Гейтс і Пол Ален) і Apple (Стів Джобс і Стефан Возняк). Утворилися співтовариства людей - хакерів, які ставлять собі за мету завдання шкоди інформаційній безпеці окремих користувачів, організацій і цілих країн. Формується інформаційне право - нова галузь міжнародної правової системи.
VII етап – починаючи з 1985 року	пов'язаний зі створенням і розвитком глобальних інформаційно-комунікаційних мереж з використанням космічних засобів забезпечення. Передбачає комплексне використання заходів і засобів захисту.
VIII етап – приблизно з кінця XX - початку XXI ст.	пов'язаний із повсюдним використанням надмобільних комунікаційних пристроїв із широким спектром завдань і глобальним охопленням у просторі та часі, що забезпечується космічними інформаційно-комунікаційними системами. Характеризується "широким переходом на цифру". Передбачає комплексне використання заходів і засобів захисту.

Комп'ютерні злочини надзвичайно багатогранні та складні явища. Об'єктами таких злочинних посягань можуть бути самі технічні засоби (комп'ютери та периферія) як матеріальні об'єкти або ж програмне забезпечення та бази даних, для яких технічні засоби є оточенням; комп'ютер може виступати як предмет посягань або як інструмент.

Види комп'ютерних злочинів надзвичайно різноманітні. Це і несанкціонований доступ до інформації, що зберігається в комп'ютері, і введення в програмне забезпечення "логічних бомб", що спрацьовують при виконанні певних умов і частково або повністю виводять з ладу комп'ютерну систему, і розробка, і розповсюдження комп'ютерних вірусів, і розкрадання комп'ютерної інформації. Комп'ютерний злочин може статися також через недбалість у розробці, виготовленні та експлуатації програмно-обчислювальних комплексів або через підробку комп'ютерної інформації [7, 8].

Серед усього набору методів захисту інформації виділяють перераховані нижче (рис. 1).

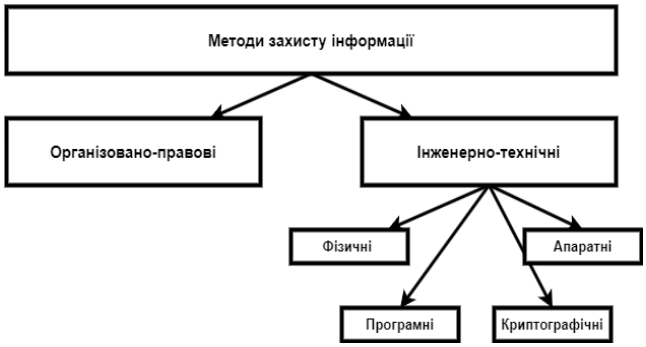


Рис. 1. Класифікація методів захисту інформації в комп'ютерних системах

Методи та засоби організаційно-правового захисту інформації. До методів і засобів організаційного захисту інформації належать організаційно-технічні та організаційно-правові заходи, що проводяться в процесі створення та експлуатації КС для забезпечення захисту інформації. Ці заходи мають проводитися під час будівництва або ремонту приміщень, у яких розміщуватимуться комп'ютери; проєктування системи, монтажу та налагодження її технічних і програмних засобів; випробувань і перевірки працездатності комп'ютерної системи.

Основою проведення організаційних заходів є використання та підготовка законодавчих і нормативних документів у сфері інформаційної безпеки, які на правовому рівні мають регулювати доступ до інформації з боку споживачів.

Методи та засоби інженерно-технічного захисту інформації. Інженерно-технічний захист (ІТЗ) - це сукупність спеціальних органів, технічних засобів і заходів щодо їх використання в інтересах захисту конфіденційної інформації.

Різноманіття цілей, завдань, об'єктів захисту і заходів, що проводяться, передбачає розгляд деякої системи класифікації засобів за видом, орієнтацією та іншими характеристиками. Наприклад, засоби інженерно-технічного захисту можна розглядати за об'єктами їхнього впливу. У цьому плані вони можуть застосовуватися для захисту людей, матеріальних засобів, фінансів, інформації. Різноманіття класифікаційних характеристик дає змогу розглядати інженерно-технічні засоби за об'єктами впливу, характером заходів, способами реалізації, масштабом охоплення, класом засобів зловмисників, яким чиниться протидія з боку служби безпеки.

За функціональним призначенням засоби інженерно-технічного захисту поділяються на такі групи:

1) *фізичні засоби*, що включають різні засоби і споруди, які перешкоджають фізичному проникненню (або доступу) зловмисників на об'єкти захисту і до матеріальних носіїв конфіденційної інформації та здійснюють захист персоналу, матеріальних засобів, фінансів та інформації від протиправних впливів;

2) *апаратні засоби* - прилади, пристрої, пристосування та інші технічні рішення, що використовуються в інтересах захисту інформації. У практиці діяльності підприємства знаходить широке застосування найрізноманітніша апаратура, починаючи з телефонного апарата до досконалих автоматизованих систем, що забезпечують виробничу діяльність. Основне завдання апаратних засобів - забезпечення стійкого захисту інформації від розголошення, витоку і несанкціонованого доступу через технічні засоби забезпечення виробничої діяльності;

3) *програмні засоби*, що охоплюють спеціальні програми, програмні комплекси та системи захисту інформації в інформаційних системах різного призначення та засобах обробки (збирання, накопичення, зберігання, обробка та передача) даних;

4) криптографічні засоби - це спеціальні математичні та алгоритмічні засоби захисту інформації, яку передають системами і мережами зв'язку, зберігають і обробляють на ЕОМ із використанням різноманітних методів шифрування.

Фізичні методи та засоби захисту інформації. Фізичні засоби захисту - це різноманітні пристрої, пристосування, конструкції, апарати, вироби, призначені для створення перешкод на шляху руху зловмисників. До фізичних засобів належать механічні, електромеханічні, електронні, електронно-оптичні, радіо- і радіотехнічні та інші пристрої для перешкоджання несанкціонованому доступу (входу, виходу), пронесенню (виносу) засобів і матеріалів, та інших можливих видів злочинних дій.

Ці засоби застосовуються для вирішення таких завдань:

1) охорона території підприємства і спостереження за нею;

2) охорона будівель, внутрішніх приміщень і контроль за ними;

3) охорона обладнання, продукції, фінансів та інформації;

4) здійснення контрольованого доступу в будівлі та приміщення.

Усі фізичні засоби захисту об'єктів можна поділити на три категорії: засоби попередження, засоби виявлення та системи ліквідації загроз. Охоронна сигналізація та охоронне телебачення, наприклад,

належать до засобів виявлення загроз; паркани навколо об'єктів - це засоби запобігання несанкціонованому проникненню на територію, а посилені двері, стіни, стелі, решітки на вікнах та інші заходи слугують захистом і від проникнення, і від інших злочинних дій (підслуховування, обстріл, кидання гранат і вибухових пакетів тощо). Засоби пожежогасіння належать до систем ліквідації загроз.

Апаратні методи та засоби захисту інформації. До апаратних засобів захисту інформації належать найрізноманітніші за принципом дії, пристроєм і можливостями технічні конструкції, що забезпечують припинення розголошення, захист від витоку і протидію несанкціонованому доступу до джерел конфіденційної інформації.

Апаратні засоби захисту інформації застосовуються для вирішення таких завдань:

1) проведення спеціальних досліджень технічних засобів забезпечення виробничої діяльності на наявність можливих каналів витоку інформації;

2) виявлення каналів витоку інформації на різних об'єктах і в приміщеннях;

3) локалізація каналів витоку інформації;

4) пошук і виявлення засобів промислового шпигунства;

5) протидія несанкціонованому доступу до джерел конфіденційної інформації та іншим діям.

Програмні методи та засоби захисту інформації

Системи захисту комп'ютера від чужого вторгнення вельми різноманітні і класифікуються, як:

1) засоби власного захисту, передбачені загальним програмним забезпеченням;

2) засоби захисту в складі обчислювальної системи;

3) засоби захисту із запитом інформації;

4) засоби активного захисту;

5) засоби пасивного захисту та інші.

Основні напрями використання програмного захисту інформації. Можна виокремити такі напрями використання програм для забезпечення безпеки конфіденційної інформації, зокрема такі:

1) захист інформації від несанкціонованого доступу;

2) захист інформації від копіювання;

3) захист програм від копіювання;

4) захист програм від вірусів;

5) захист інформації від вірусів;

6) програмний захист каналів зв'язку.

За кожним із зазначених напрямів є достатня кількість якісних, розроблених професійними організаціями і розповсюджуваних на ринках програмних продуктів.

Програмні засоби захисту мають такі різновиди спеціальних програм:

1) ідентифікації технічних засобів, файлів та автентифікації користувачів;

2) реєстрації та контролю роботи технічних засобів і користувачів;

3) обслуговування режимів обробки інформації обмеженого користування;

4) захисту операційних засобів ЕОМ і прикладних програм користувачів;

5) знищення інформації в захисні пристрої після використання;

6) сигналізують порушення використання ресурсів;

7) допоміжних програм захисту різного призначення.

Захист інформації від несанкціонованого доступу

Для захисту від чужого вторгнення обов'язково передбачаються певні заходи безпеки. Основні функції, які мають здійснюватися програмними засобами, це:

1) ідентифікація суб'єктів та об'єктів;

2) розмежування (іноді й повна ізоляція) доступу до обчислювальних ресурсів та інформації;

3) контроль і реєстрація дій з інформацією та програмами.

Найпоширенішим методом ідентифікації є парольна ідентифікація. Однак практика показує, що парольний захист даних є слабкою ланкою, оскільки пароль можна підслухати або підглянути, перехопити або просто розгадати.

Захист від копіювання. Засоби захисту від копіювання запобігають використанню крадених копій програмного забезпечення і є нині єдиним надійним засобом, який як захищає авторське право програмістів-розробників, так і стимулює розвиток ринку. Під засобами захисту від копіювання розуміють засоби, що забезпечують виконання програмою своїх функцій тільки в разі розпізнання деякого унікального елемента, що не копіюється. Таким елементом (званим ключовим) може бути дискета, певна частина комп'ютера або спеціальний пристрій, який під'єднують до персонального комп'ютера.

Захист від копіювання реалізується виконанням низки функцій, які є загальними для всіх систем захисту:

1) ідентифікація середовища, з якого запускатиметься програма (дискета або ПК);

2) аутентифікація середовища, з якого запущено програму;

3) реакція на запуск із несанкціонованого середовища;

4) реєстрація санкціонованого копіювання;

5) протидія вивченню алгоритмів роботи системи.

6) захист програм і даних від комп'ютерних вірусів

Шкідницькі програми і, насамперед, віруси становлять дуже серйозну небезпеку при зберіганні на ПЕОМ конфіденційної інформації. Недооцінка цієї небезпеки може мати серйозні наслідки для інформації користувачів. Знання механізмів дії вірусів, методів і засобів боротьби з ними дає змогу ефективно організувати протидію вірусам, звести до мінімуму ймовірність зараження і втрат від їхнього впливу.

"Комп'ютерні віруси" - це невеликі виконувані або інтерпретовані програми, що мають властивість розповсюдження і самовідтворення (реплікації) в комп'ютерній системі. Віруси можуть виконувати зміну або знищення програмного забезпечення або

даних, що зберігаються в ПЕОМ. У процесі поширення віруси можуть себе модифікувати.

Класифікація комп'ютерних вірусів. Нині у світі налічується понад 40 тисяч тільки зареєстрованих комп'ютерних вірусів. Оскільки переважна більшість сучасних шкідницьких програм мають здатність до саморозмноження, то часто їх відносять до комп'ютерних вірусів.

Усі комп'ютерні віруси можуть бути класифіковані за такими ознаками:

- за середовищем існування вірусу,
- за способом зараження середовища проживання,
- за деструктивними можливостями,
- за особливостями алгоритму вірусу.

Масове поширення вірусів, серйозність наслідків їхнього впливу на ресурси комп'ютерів спричинили необхідність розроблення та використання спеціальних антивірусних засобів і методів їхнього застосування. Антивірусні засоби застосовуються для вирішення таких завдань:

- виявлення вірусів у КС,
- блокування роботи програм-вірусів,
- усунення наслідків впливу вірусів.

Виявлення вірусів бажано здійснювати на стадії їх впровадження або, принаймні, до початку здійснення деструктивних функцій вірусів. Необхідно зазначити, що не існує антивірусних засобів, які гарантують виявлення всіх можливих вірусів.

У разі виявлення вірусу необхідно одразу ж припинити роботу програми-вірусу, щоб мінімізувати збитки від його впливу на систему.

Усунення наслідків впливу вірусів ведеться у двох напрямках:

- видалення вірусів,
- відновлення (за необхідності) файлів, областей пам'яті.

Для боротьби з вірусами використовуються програмні та апаратно-програмні засоби, які застосовуються в певній послідовності та комбінації, утворюючи методи боротьби з вірусами.

Найнадійнішим методом захисту від вірусів є використання апаратно-програмних антивірусних засобів. Нині для захисту ПК використовуються спеціальні контролери та їх програмне забезпечення. Контролер встановлюється в роз'єм розширення і має доступ до загальної шини. Це дає йому змогу контролювати всі звернення до дискової системи. У програмному забезпеченні контролера запам'ятовуються області на дисках, зміна яких у звичайних режимах роботи не допускається. Таким чином, можна встановити захист на зміну головного завантажувального запису, завантажувальних секторів, файлів конфігурації, виконуваних файлів тощо.

У разі виконання заборонених дій будь-якою програмою контролер видає відповідне повідомлення користувачеві і блокує роботу ПК.

Апаратно-програмні антивірусні засоби мають низку переваг перед програмними:

- працюють постійно;
- виявляють усі віруси, незалежно від механізму їхньої дії;
- блокують недозволені дії, які є результатом роботи вірусу або некваліфікованого користувача.

Недолік у цих засобів один - залежність від апаратних засобів ПЕОМ. Зміна останніх веде до необхідності заміни контролера.

Сучасні програмні антивірусні засоби можуть здійснювати комплексну перевірку комп'ютера на предмет виявлення комп'ютерних вірусів. Для цього використовуються такі антивірусні програми як - Kaspersky Anti-Virus (AVP), Norton Antivirus, Dr. Web, Symantec Antivirus. Усі вони мають антивірусні бази, які періодично оновлюються.

Криптографічні методи та засоби захисту інформації. Криптографія як засіб захисту (закриття) інформації набуває дедалі важливішого значення у світі комерційної діяльності.

Криптографія має досить давню історію. Спочатку вона застосовувалася головним чином у сфері військового і дипломатичного зв'язку. Тепер вона необхідна у виробничій і комерційній діяльності. Якщо врахувати, що сьогодні каналами шифрованого зв'язку тільки у нас у країні передають сотні мільйонів повідомлень, телефонних переговорів, величезні обсяги комп'ютерних і телеметричних даних, і все це не для чужих очей і вух, стає зрозумілим: збереження таємниці цієї інформації тут украй необхідне.

Криптографія охоплює кілька розділів сучасної математики, а також спеціальні галузі фізики, радіоелектроніки, зв'язку та деяких інших суміжних галузей. Її завданням є перетворення математичними методами переданого каналами зв'язку секретного повідомлення, телефонної розмови або комп'ютерних даних таким чином, що вони стають абсолютно незрозумілими для сторонніх осіб. Тобто криптографія повинна забезпечити такий захист секретної (або будь-якої іншої) інформації, що навіть у разі її перехоплення сторонніми особами та обробки будь-якими способами з використанням найшвидкодіючіших ЕОМ і останніх досягнень науки і техніки, вона не має бути дешифрована протягом кількох десятків років. Для такого перетворення інформації використовують різні шифрувальні засоби - такі, як засоби шифрування документів, зокрема й портативного виконання, засоби шифрування мовлення (телефонних і радіопереговорів), телеграфних повідомлень і передавання даних.

Загальна технологія шифрування. Початкова інформація, яка передається каналами зв'язку, може являти собою мову, дані, відеосигнали, називається незашифрованими повідомленнями Р.

У пристрої шифрування повідомлення Р шифрується (перетворюється на повідомлення С) і передається "незакритим" каналом зв'язку. На приймальній стороні повідомлення С дешифрується для відновлення вихідного значення повідомлення Р. Параметр, який може бути застосований для вилучення окремої інформації, називається ключем. Якщо в процесі обміну інформацією для шифрування і читання використовувати один той самий ключ, то такий криптографічний процес називається симетричним. Його основним недоліком є те, що перш

ніж почати обмін інформацією, потрібно виконати передачу ключа, а для цього необхідний захищений зв'язок.

Нині під час обміну даними каналами зв'язку використовується несиметричне криптографічне шифрування, засноване на використанні двох ключів. Це нові криптографічні алгоритми з відкритим ключем, засновані на використанні ключів двох типів: секретного (закритого) і відкритого.

У криптографії з відкритим ключем є принаймні два ключі, один з яких неможливо обчислити з іншого. Якщо ключ розшифрування обчислювальними методами неможливо отримати з ключа зашифрування, то секретність інформації, зашифрованої за допомогою несекретного (відкритого) ключа, буде забезпечена. Однак цей ключ має бути захищений від підміни або модифікації. Ключ розшифрування також має бути секретним і захищений від підміни або модифікації.

Якщо, навпаки, обчислювальними методами неможливо отримати ключ зашифрування з ключа розшифрування, то ключ розшифрування може бути не секретним.

Ключі влаштовані таким чином, що повідомлення, зашифроване однією половинкою, можна розшифрувати тільки іншою половинкою. Створивши пару ключів, компанія широко поширює відкритий (публічний) ключ і надійно охороняє закритий (особистий) ключ.

Захист публічним ключем не є абсолютно надійним. Вивчивши алгоритм його побудови, можна реконструювати закритий ключ. Однак знання алгоритму ще не означає можливість провести реконструкцію ключа в розумно прийнятні терміни. Виходячи з цього, формується принцип достатності захисту інформації: захист інформації прийнято вважати достатнім, якщо витрати на його подолання перевищують очікувану вартість самої інформації. Цим принципом керуються під час несиметричного шифрування даних.

Поділ функцій зашифрування і розшифрування за допомогою поділу на дві частини додаткової інформації, необхідної для виконання операцій, є тією цінною ідеєю, яка лежить в основі криптографії з відкритим ключем.

Криптографічному захисту фахівці приділяють особливу увагу, вважаючи його найнадійнішим, а для інформації, що передається по лінії зв'язку великої протяжності, – єдиним засобом захисту від розкрадань.

Висновки

1. Поняття інформації ємне, багатогранне, а її визначення багато в чому залежить від галузі людської діяльності.
2. Інформація, як об'єктивне відображення реальності, може існувати в різних формах і мати певні властивості, водночас людина може сприймати за допомогою органів чуття 5 її видів.
3. Поняття інформаційної безпеки є ключовою умовою успіху виробничої та підприємницької діяльності, і включає в себе питання інформаційно-комерційної, юридичної та фізичної безпеки.
4. Класифікація методів захисту інформації включає в себе організаційно-правові, інженерно-технічні методи, які, своєю чергою, складаються з фізичних, апаратних, програмних і криптографічних методів.

СПИСОК ЛІТЕРАТУРИ

1. Безруков, В.В. Теорія інформації / В.В. Безруков, В.Я. Кізяков, В.І. Профатілов. – Д.: ДДТУЗТ, 2001. – 110 с.
2. Жураковський Ю.П., Полторак В.П. Теорія інформації та кодування: Підручник.- К.:Вища школа, 2011. — 255 с.
3. Тулякова Н. О. Теорія інформації: Навчальний посібник. - Суми: Вид-во СумДУ, 2008.- 212 с.
4. Кулик А.Я., Кривогубченко С.Г. Теорія інформації і кодування / Навч. посібник. – Вінниця: ВНТУ, 2008. 145 с.
5. Ali Al-Ammouri. Development of a mathematical model of reliable structures of information-control systems / Ali Al-Ammouri, Iryna Lebid, Marina Dekhtiar, Ievgenii Lebid, Hasan Al-Ammori // Eastern-European Journal of Enterprise Technologies. – 2022. – Vol. 5/9, Issue (119).– P. 68–78. DOI: https://doi.org/10.15587/1729-4061.2022.265953.
6. Інформаційні системи та мережі: навчальний посібник. / Аль-Амморі. А.Н, Лясковський В.П., Попова Л.С., Тимченко О.П, Полєва Н.М. – К-НТУ-2021, 194с.
7. Подлевський Б. М. Теорія інформації : підручник / Б. М. Подлевський, Р. Є. Рикалюк. – Львів: Видавничий центр ЛНУ ім. І. Франка, 2016. – 342 с.
8. Методологія і технології захисту інформації: навчальний посібник / А.Н. Аль-Амморі, Н.М. Наумова, П.В. Дяченко, Р.М. Іщенко, М.М. Дехтяр, А.Є. Клочан; НТУ. – Київ: НТУ, 2020. – 147с.

Received (Надійшла) 22.11.2023
Accepted for publication (Прийнята до друку) 10.01.2024

Methods and means of protecting information

A. Al-Ammouri, M. Dekhtyar, R. Ishchenko, E. Klochan

Abstract. The article discusses general issues of organizing methods and means of information protection. Various definitions of the general scientific concept of "information" are considered, from the point of view of various scientists, researchers, and depending on the branch of human activity. The types of information presentation and its individual properties in relation to computer data processing are considered. The fundamental concepts and definitions from the field of information security of systems are considered. The historical stages in the development of information security means are given, the classification of information security methods is given, the main directions of their use are investigated. The classification of computer viruses by their main features, as well as the tasks solved by antivirus tools, are considered. Cryptographic methods of information protection and general encryption technology are considered separately.

Keywords: information, information security, confidentiality, integrity, availability.

УДК 004.825 **doi: 10.26906/SUNZ.2024.1.045**

М. Р. Братищенко, Т. В. Філімончук, Г. В. Майстренко, В. І. Сітніков

Харківський національний університет радіоелектроніки, Харків, Україна

АНАЛІЗ МЕТОДІВ ВИЯВЛЕННЯ АНОМАЛІЙ У ДАНИХ ПРО СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ

Анотація. Актуальність. Сучасні люди постійно шукають нові способи використання енергії для покращення свого життя, тому попит на неї зростає. У більшості випадків компаніям і галузям важко контролювати всі свої пристрої одночасно, що може призвести до втрати електроенергії в будь-який час. В результаті операційні витрати будуть більшими, ніж необхідно. Крім того, втрата електроенергії сприяє глобальному потеплінню через вивільнення вуглецю, коли енергія генерується шляхом спалювання вугілля, газу та нафти. Отже, потрібні рішення для вирішення цих проблем. **Метою даної роботи** є аналіз існуючих методів виявлення аномалій в даних задля вирішення проблеми надмірного споживання електроенергії та попередження про критичні значення в показниках електроенергії, споживаної різноманітними пристроями та електрообладнанням. **Об'єктом дослідження** є процес виявлення нетипових значень або значних відхилень в показниках споживання електроенергії за такими параметрами, як: напруга, сила, частота струму, потужність**. Предметом дослідження** є моделі та методи виявлення аномалій в даних. **Результати.** Після ретельного аналізу кожного з перелічених методів виявлення аномалій відкриваються нові можливості для вирішення проблеми енергоспоживання. Наприклад: об'єднання декількох методів в один; розробка моделі машинного навчання на основі одного або декількох методів, тренування на тестових даних і в перспективі оброблення реальних даних про енергоспоживання з метою визначення нетипових значень, з можливістю фіксування дати та часу виникнення аномалій, а також побудовою різноманітних графіків на основі цієї інформації. **Висновок.** Завдяки розглянутим методам виявлення аномалій можна запобігти великому споживанню електроенергії для досягнення енергозбереження, нагадувати користувачам про визначення несправних електроприладів або змінювати неправильні схеми споживання електроенергії, знижувати витрати користувачів на енергоспоживання та сприяти обізнаності щодо безпеки споживання електроенергії.

Ключові слова: виявлення аномалій, енергоспоживання, статистичні дані, машинне навчання, кластеризація, методи.

Вступ

На сьогоднішній день людство використовує електроенергію багато в яких сферах свого життя, починаючи від побутових електроприладів, і закінчуючи електрообладнанням на підприємствах та електростанціях. Проте треба пам'ятати, що надмірне споживання електроенергії це погано, а особливо якщо є різкі скачки напруги – можуть погоріти електроприлади або електрообладнання вийде з ладу. Виявлення аномалій може запобігти великому споживанню електроенергії для досягнення енергозбереження, нагадувати користувачам про виявлення несправних електроприладів або змінювати неправильні схеми споживання електроенергії, знижувати витрати користувачів на енергоспоживання та сприяти обізнаності щодо безпеки споживання електроенергії.

На початку аналізу досліджень стосовно існуючих методів та рішень щодо врегулювання енергоспоживання та ідентифікації нетипових значень варто ознайомитись із статистичними даними про енергоспоживання у різних країнах світу за останні роки.

У статті [1] представлені та проаналізовані дані, взяті з кількох досліджень про енергоспоживання будівель у США, ЄС та БРІК країн (Бразилія, Росія, Індія, Китай). Більшість поточних досліджень споживання енергії стосується статистики конкретної країни. Однак міжнародні порівняння корисні для виявлення історичних, фактичних тенденцій та тенденцій споживання енергії. Дані представлені у звітах Світового банку, Програми ООН з навколишнього середовища, Міжурядової групи експертів зі змін клімату та міжнародного енергетичного агентства. Вони порівнюються з національними звітами, а також із дослідженнями. Цей аналіз показує, що країни БРІК вже подолали загальне енергоспоживання розвинутих країн. Але розширення їх будівельного фонду викликає нагальну потребу в енергоефективності в будівлях. Водночас, можна зробити висновок, що заходів, прийнятих у розвинених країнах, недостатньо для того, щоб гарантувати значне скорочення енергії споживання в будівлях та на підприємствах.

У статті [2] проаналізовано виявлення аномалій у даних щодо споживання електроенергії за допомогою 2 методів: Isolation Forest та Gaussian Naïve Bayes. Автори докладно описують сутності цих методів, наводять формули для розрахунку ймовірностей виникнення аномалій у даних за період часу і пропонують моделі, які можливо натренувати на основі цих методів. Результатом стали аналіз цих даних споживання електроенергії з відповідними графіками. Проведене моделювання показало, що було досягнуто збалансованої оцінки точності щонайменше 0,8947.

Підхід Isolation Forest використовувався для позначення нормальних та ненормальних даних, і він успішно ідентифікував аномальні стрибки на основі поведінки моделі споживання енергії, а не просто на величині споживання. У виявленні аномалій Gaussian Naïve Bayes дав задовільну продуктивність у визначенні ненормальних, а також нормальних точок. Прогнозування споживання електроенергії та ідентифікація аномалій є критично важливими у функціонуванні енергомережі, а обробка багатозмінних часових рядів є складною задачею. Автори статті [3] представили модель, яка поєднує підходи Transformer та K-means. Кожні 23 години навчальні дані розділяються на k кластерів за допомогою кластеризації K-середніх. У той же час ці навчальні дані використовуються для навчання

моделі Transformer для прогнозування споживання електроенергії за наступну годину, при цьому прогнозоване значення поміщається в навчений кластер K-середніх, а центроїд кластера виступає як остаточне прогнозоване значення. Нарешті, для визначення аномалії, було зроблене порівняння очікуваного значення з фактичними результатами тесту. Експериментальні результати у вигляді графіків доводять, що модель забезпечує точність прогнозування з меншою похибкою та високою ефективністю виявлення аномалій.

В дослідженні [4] автори охопили багато питань, починаючи від теоретичних відомостей щодо аномалій в даних та їх виявлення, аналіз існуючих математичних методів та алгоритмів машинного навчання, і закінчуючи розробкою концепції фреймворку для виявлення аномалій споживання електроенергії за допомогою хмарних обчислень. На початку було класифіковано 3 види аномалій: точкові, контекстуальні та колективні. Далі проведено аналіз методів для виявлення аномалій, які найбільш підходять під кожен вид. На основі аналізу автори запропонували своє бачення, яким чином можна визначати нетипові показники у даних про споживання електроенергії за допомогою хмарних обчислень. Наприкінці дослідження було приділено увагу точності виявлення аномалій та можливі способи покращення цього показника.

Автори публікації [5] застосували нову техніку лямбда-архітектури до системи виявлення аномалій у даних споживанні електроенергії, щоб підтримувати пакетне оновлення моделей та моніторинг в реальному часі. Вони запропонували алгоритм викриття для пошуку аномалій на основі історії споживання за допомогою контрольованого навчання та статистичних алгоритмів. Крім того, система підтримує персоналізовану службу оповіщення шляхом встановлення порогового значення для ненормальних показників споживання енергії. Було оцінено точність виявлення аномалії алгоритму на наборі даних реального світу та масштабованість системи на великому наборі синтетичних даних. Результати підтвердили ефективність запропонованої системи з лямбда-архітектурою.

Метою роботи є аналіз існуючих методів виявлення аномалій в даних задля вирішення проблеми надмірного споживання електроенергії та попередження про критичні значення в показниках.

Основна частина

Виявлення аномалій – це метод розпізнавання даних, які відрізняються від звичайних. Аномалії в даних – це ситуації, які не відповідають визначеній звичайній моделі поведінки. Для того щоб проаналізувати та ідентифікувати аномалії треба сформувати набір даних за певний період часу (або використовувати вже готові дані). Також можливий аналіз даних в реальному часі. Аномалії, загалом, також відомі як викиди, девіанти, неузгодженості або винятки. Як правило, аномалії бувають наступних типів:

- точкові аномалії, які є найпростішим і дуже поширеним випадком. Точкові аномалії часто представляють екстремум, нерегулярність або відхилення, що трапляються випадковим чином і не мають особливого значення;

- контекстуальні аномалії, які часто виявляються в часових рядах та просторових даних. Це випадки, які можна розглядати як аномальні у якомусь конкретному контексті. Але варто зазначити, що спостереження однієї і тієї ж точки в різних контекстах не завжди дасть ознаки аномальної поведінки;

- колективні аномалії – це група корельованих, взаємопов'язаних або послідовних випадків, коли кожен конкретний екземпляр сам по собі не повинен бути аномальним, але їх колективне виникнення є аномальним.

Існують три основні категорії методів виявлення аномалій: контрольована, напівконтрольована та без нагляду.

Методи виявлення аномалій без нагляду (неконтрольовані алгоритми) визначають аномалії на непозначеному наборі даних, виходячи з припущення, що більшість зразків у цьому набору є нормальними, і шукаючи зразки, що виглядають якнайменше відповідними решті набору даних.

Методи контрольованого виявлення аномалій вимагають набору даних, що позначено як «нормальні» або «аномальні», та включають навчання класифікатора (ключовою відмінністю від інших задач класифікації є притаманно незбалансований характер виявлення викидів).

Методи напівконтрольованого виявлення аномалій створюють модель, що представляє нормальну поведінку, виходячи із заданого нормального навчального набору даних, і потім перевіряють правдоподібність того, що тестовий екземпляр було породжено вивченою моделлю.

Метод Isolation Forest. Isolation Forest – це один із алгоритмів неконтрольованого машинного навчання, який використовується для виявлення аномалій у наборі даних [2]. На відміну від керованих алгоритмів машинного навчання, Isolation Forest не потребує жодних міток чи класифікації для даних, які потрібно проаналізувати. Алгоритм або відокремлює аномалії, розглядаючи аномалії як випадки, які менш імовірні, або приписує значення, які дуже відрізняються від зазвичай приписуваних. На рис. 1 можна побачити алгоритм Isolation Forest для визначення аномалій в даних щодо споживання електроенергії. Класифіковані значення 0 використовувалися для позначення нормального споживання енергії, тоді як 1 вказувало на аномальне споживання енергії.

Метод Clustering of K-means. K-Means – це один із алгоритмів, що використовується для кластеризації та розбиває дані на кілька груп. Алгоритм K-Means є одним із методів неієрархічної кластеризації даних, який може групувати дані в кілька кластерів на основі подібності даних [6]: дані з однаковими характеристиками в один кластер, а дані, які мають різні характеристики, групуються в інші кластери. Щоб визначити мітку кластера будь-яких даних, обчислюється відстань між даними з кожним кластером центр.

Є кілька способів, якими можна скористатися для обчислення відстані, наприклад Евклідова відстань, відстань Манхеттен і відстань Чебічі. Метод K-Means спрямований на мінімізацію суми квадратів відстаней між усіма точками та центром кластера.

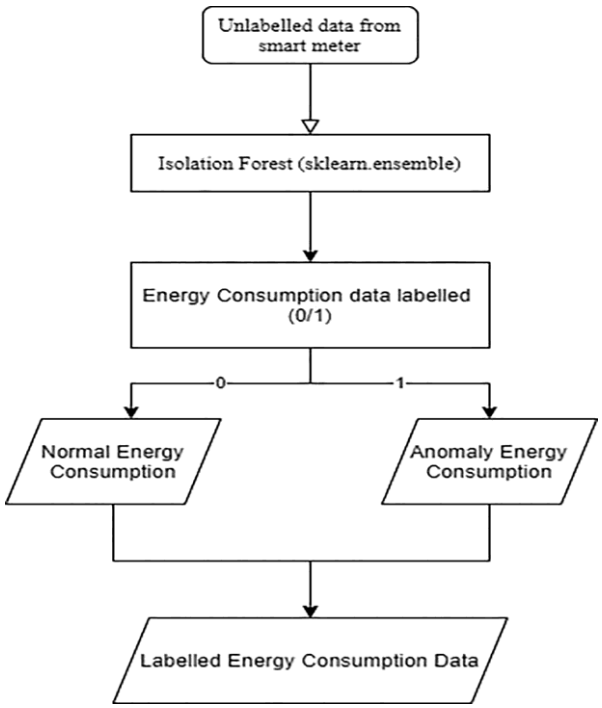


Рис. 1. Блок-схема маркування даних за алгоритмом Isolation Forest

Ця процедура складається з наступних кроків.

Крок 1: обираємо k із заданих n шаблонів як початкові центри кластерів. Призначаємо кожний шаблон, що залишився, до одного з k кластерів; шаблон призначається найближчому центру/кластеру.

Крок 2: обчислюємо центри кластерів на основі поточного призначення шаблонів.

Крок 3: відносимо кожен із n шаблонів до їх найближчого центру/кластеру.

Крок 4: якщо немає змін у призначенні шаблонів кластерам протягом двох послідовних ітерації, завершуємо процедуру, інакше повертаємось до кроку 2.

В [6] автори взяли тестові дані з показниками споживання електроенергії одним підприємством на протязі 1 року і розділили на 4 частини (на кожну пору року). На цих даних і було проведено тренування та апробація результатів кластеризації. Для кожного сценарію було виміряно сумарну квадратичну помилку (SSE) та кількість ітерацій. SSE описує значення стандартного відхилення кожного кластера до центру обробки даних. Більше значення SSE означає, що ступінь подібності даних в одному кластері нижче. Кількість ітерацій описує довжину кластерів на протязі процесу формування. Результати кластеризації наведено в табл. 1.

Метод Support Vector Machine. Support vector machine (SVM) є керованим алгоритмом машинного навчання, який часто використовують для класифікації. SVM використовує гіперплощини в багатовимірному просторі, щоб розділити точки даних на класи. SVM зазвичай застосовується, коли в проблему залучено більше ніж один клас. Однак у виявленні аномалій він також використовується для проблем одного класу. Модель натренована визначати «норму» і може зрозуміти, чи належать незнайомі дані до цього класу, чи являють собою аномалію.

Таблиця 1 – **Результати кластеризації**

№	Сценарій	SSE	Кількість ітерацій
1	4 кластери без усунення аномалій	0.174	20
2	4 кластери з усуненням аномалій	0.752	14
3	5 кластерів без усунення аномалій	0.134	21
4	5 кластерів з усуненням аномалій	0.509	22

Багато факторів сприяли високій популярності SVM сьогодні. Наприклад: рішення розріджене, що робить його дійсно ефективним у порівнянні з іншими підходами на основі ядра. Також він може використовувати нелінійне перетворення в формі ядра, яке навіть дозволяє розглядати SVM як техніку зменшення розмірності [7]. Однокласовий SVM розроблений для випадків, коли відомий лише один клас і треба виявляти будь-що поза цим класом. Це відомо як визначення новизни і відноситься до автоматичної ідентифікації непередбачених або аномальних явищ, тобто викиди, вбудовані у велику кількість нормальних даних. На відміну від традиційного SVM, One-Class SVM (OC-SVM) дозволяє визначити межу, яка досягає максимальне розділення між зразками «відомий клас» та «походження». Лише невелика частина точок даних може лежати з іншого боку межі прийняття рішення: ці точки даних вважаються викидами (рис. 2).

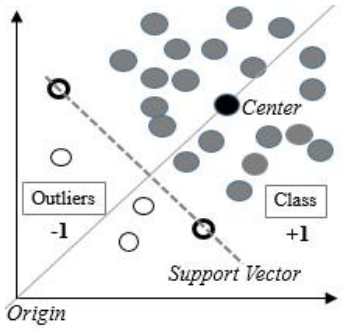


Рис. 2. Класифікація точок відносно опорного вектора класу

Метод Local Outlier Factor. Алгоритм LOF – це неконтрольований метод виявлення викидів, який оцінює унікальність кожної події на основі відстані від k-найближчих сусідів [8]. Алгоритм LOF здатний виявляти викиди незалежно від розподілу даних, оскільки він робить певні припущення щодо їх розподілу. Основна ідея алгоритму полягає в тому, що щільність навколо стороннього об'єкта суттєво відрізняється від щільності навколо своїх сусідів. Це перевага, коли дані, які аналізуються, не позначені або неможливо позначити через великий обсяг даних. Такий варіант поширено в комп'ютерних мережах де ряд генерованих мережних пакетів дуже високий.

На рис. 3 наведено схему обробки даних та виявлення аномалій. Перший крок складається з навчання NSL-KDD та розділення наборів даних на звичайні та атакуючі. Тільки для навчання використовуються записи, що відповідають звичайним даним. Далі, необхідно видалити атрибути 8 та 20 зі звичай-

ного набору даних, оскільки всі значення цих атрибутів дорівнюють 0, тобто вони не мають передбачуваної сили. Наступний крок це переведення даних на стандартизовану Z шкалу, де середнє значення буде дорівнювати 0, а стандартне відхилення 1. Це необхідно задля визначення допустимих значень та відхилень (аномалій). Даний процес застосовується до всіх числових значень і номінальних атрибутів (2 – тип протоколу, 3 – сервіс, 4 – прапор). Після обробки отримуємо 75 атрибутів. Перед застосуванням алгоритму LOF необхідно вказати два параметри: кількість найближчих сусідів k та граничне значення, щоб виявляти чи запис виходить за межі чи ні. При виборі кількість найближчих сусідів, рекомендується щоб k-значення дорівнюватиме квадратному кореню з усіх даних, які використано для модельного навчання. За алгоритмом LOF записи, які мають граничне значення відхилення більше 1, вважаються викидами. Для нормальної обробки даних існує шість порогових значень: Thc (cleaning) = {1,5; 1,75; 2; 3; 5 та 10}.

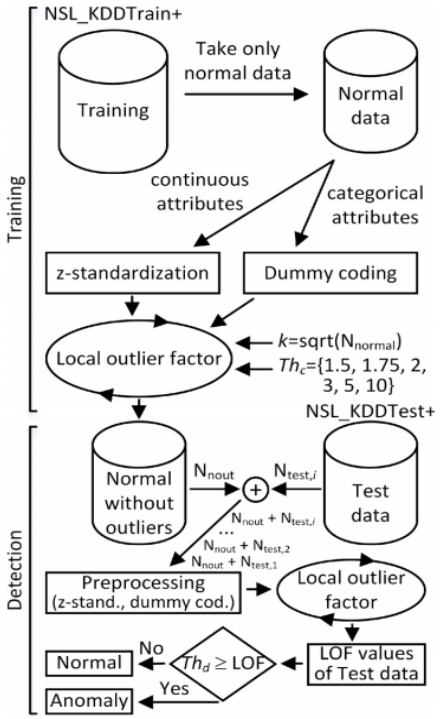


Рис. 3. Обробка даних та виявлення аномалій в LOF

Алгоритм LOF застосовується з обраною кількістю найближчих сусідів k та граничним значенням. На основі розрахованих значень, записи зі значеннями граничного відхилення, які вище або дорівнюють Thc видаляються. Тоді k-значення для нормального типу даних, з якого були викиди видаляються, перераховуються, і знову виконується алгоритм LOF. Цикл повторюється до тих пір, поки не залишиться жодних записів, які будуть перевищювати встановлене граничне значення. Після навчання, отриманий набір даних далі використовується для виявлення аномалій.

Після фази навчання було підготовлено шість наборів даних, які складаються з даних нормального типу з видаленими аномаліями. Кількість викидів, знайдених та видалених за допомогою обраних порогових значень наведено на рис. 4. Числа в дужках – це відсотки записів, які було видалено з набору навчальних даних.

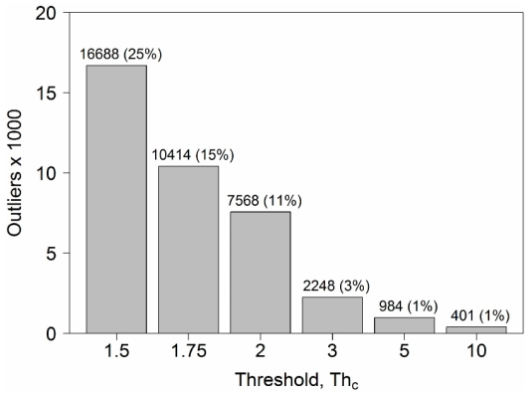


Рис. 4. Кількість аномалій, які було видалено з нормальних даних використовуючи значення Thc

З рисунку видно, що перші три набори даних із пороговими значеннями 1.5, 1.75 та 2.0 повинні мати найбільший вплив на результати виявлення, оскільки ці набори даних зменшуються на 25, 15 та 11 відсотків відповідно.

Метод Unsupervised Niche Clustering. Unsupervised Niche Clustering (UNC) – це надійний метод кластеризації, який використовує еволюційний алгоритм зі стратегією заняття ніш [9]. Даний алгоритм допомагає знаходити кластери за допомогою стійкої функції пристосованості, в той час як техніка заняття ніш дозволяє створювати та підтримувати ніші (кластери-кандидати). Оскільки UNC базується на генетичній оптимізації, він набагато менш сприйнятливий до неоптимальних рішень, ніж традиційні методи. Основною перевагою алгоритму є здатність обробляти шум і автоматично визначати кількість кластерів. Автори статті [10] поєднали UNC з теорією нечітких множин для виявлення аномалій і застосували його для ідентифікації мережуих вторгнень. Вони пов'язані з кожним кластером, згенерованим за допомогою UNC. Це функції-члени, які відповідають гаусовій формі, використовуючи еволюційний центр та радіус кластера. Такі функції приналежності кластерів визначатимуть рівень нормалізації вибірки даних.

Метод Regression Model-Based. Виявлення аномалій на основі регресійних моделей є підкатегорією параметричних методів [11], що включає низку методів, які широко застосовуються до даних часових рядів. Ці методи базуються на двоетапному підході. Спочатку на навчальних даних будується регресійна модель. Потім отримана модель використовується на тестових послідовностях для обчислення залишків, наприклад, різниці між прогнозованим значенням і реальним значенням. На основі залишків остаточно визначаються оцінки аномалій. До цієї категорії можна віднести методи виявлення аномалій, які засновані на традиційних моделях прогнозування часових рядів, таких як векторна авторегресія (VAR) [12] та авторегресійне інтегроване ковзне середнє (ARIMA).

Висновки

В результаті проведених авторами досліджень було зроблено аналіз популярних існуючих рішень для визначення аномалій в показниках щодо споживання електроенергії. При чому дослідження було проведено як з метою теоретичного ознайомлення з принципом роботи того чи іншого методу, так і експериментальною перевіркою тестових даних задля визначення коректності та точності виявлення аномалій. Звернута увага на обмеження даних методів обробки даних, а також на факти, висновки, рекомендації, закономірності з раніше відомих досліджень.

Після ретельного огляду кожного методу можна зробити висновок, що кожен з них підходить для вирішення поставленої задачі – потрібно лише розробити модель на основі обраного методу машинного навчання (або об'єднати декілька методів). Потім провести «тренування» на тестових даних, після чого можна аналізувати вже реальні дані за певний проміжок часу про енергоспоживання за такими критеріями як напруга (В), сила струму (А), потужність (Вт), частота (Гц). Було б доцільно не тільки обробляти дані та визначати нетипові значення, а ще і формувати це у вигляді графіків задля наочного бачення аномальних значень. В ідеальному сценарії результатом досліджень є розробка моделі, яка буде обробляти ці дані в реальному часі.

З точки зору як теоретичної, так і прикладної перевірки цей аналіз має вагоме значення. Адже тема енергетики була, є і буде актуальною. Тим паче попередження про різкі показники напруги дозволять попередити вихід зі строю електроприладів та завчасно їх вимкнути. Або у випадку збору статистичних даних це дасть змогу зрозуміти, коли і як часто показники мали пікові значення, що дасть розуміння щодо подальшого обслуговування та експлуатації електрообладнання.

СПИСОК ЛІТЕРАТУРИ

1. Berardi U. (2015), "Building energy consumption in US, EU, and BRIC countries", Department of Architectural Science, Faculty of Engineering and Architectural Science, № 118, P. 128-136.
2. Jia Yan Lim, Wooi-Nee Tan, Yi-Fei Tan (2022), "Anomalous energy consumption detection using a Naïve Bayes approach", Faculty of Engineering, Multimedia University, Cyberjaya, Selangor, 63100, Malaysia, № 1, P. 4.
3. Zhang J., Zhang H., Ding S., Znang X. (2021), "Power Consumption Predicting and Anomaly Detection Based on Transformer and K-Means", College of Mathematics and Inf. Technology, Hebei University. Vol. 9, Article № 779587, P. 3-7.
4. Longji Feng, Shu Xu, Linghao Zhang, Jing Wu, Jidong Zhang, Chengbo Chu, Zhenyu Wang and Haoyang Shi (2020), "Anomaly detection for electricity consumption in cloud computing: framework, methods, applications, and challenges", *Wireless Com Network*, № 194, P. 2-10, doi: https://doi.org/10.1186/s13638-020-01807-0
5. Liu X., Iftikhar N., Nielsen P.S., Heller A. (2016), "Online Anomaly Energy Consumption Detection Using Lambda Architecture", Big Data Analytics and Knowledge Discovery, vol 9829, P. 193-209, doi: https://doi.org/10.1007/978-3-319-43946-4_13
6. Yasirli Amril, Amanda Lailatul Fadhilah, Fatmawati, Novi Setianil, Septia Ranil (2016), "Analysis Clustering of Electricity Usage Profile Using K-Means Algorithm", IOP Conf. Series. № 105, P.2-7, doi: 10.1088/1757-899X/105/1/012020
7. Lamrini B., Gjini A., Daudin S., Armando F., Pratmarty P., Travé-Massuyès L. (2018) "Anomaly Detection Using Similarity-based One-Class SVM for Network Traffic Characterization", Un-té de Toulouse, CNRS, Toulouse, France., № 1, P. 2-4.
8. Auskalnis J., Paulauskas N., Baskys A. (2018), "Application of Local Outlier Factor Algorithm to Detect Anomalies in Computer Network", Elektronika I Elektrotechnika, 24(3), P. 96-99, doi: https://doi.org/10.5755/j01.eie.24.3.20972
9. Nasraoui O., Leon E., Krishnapuram R. (2005), "Unsupervised Niche Clustering: Discovering an Unknown Number of Clusters in Noisy Data Sets"., Evolutionary Computation in Data Mining, Springer Berlin Heidelberg. Volume 1, P. 2-30.
10. Lizabeth Leon., Olfa Nasraoui., Jonatan Gomez. (2007), "Anomaly detection based on unsupervised niche clustering with application to network intrusion detection", Proc. of the IEEE Conference on Evolutionary Computation. Volume 1, P. 502.
11. Chandola V., Banerjee A., Kumar V. (2009), "Anomaly detection: A survey", ACM Computing Surveys, № 41, P. 1-58.
12. Melnyk I., Matthews B., Valizadegan H., Banerjee A., Oza N. (2016), "Vector Autoregressive Model-Based Anomaly Detection in Aviation Systems", JAIS Aerospace Information Systems., 13, P. 161-173.

Received (Надійшла) 22.12.2023
(Accepted for publication) Прийнята до друку 17.01.2024

Analysis of methods for detecting anomalies in electricity consumption data

Mykyta Bratyshchenko, Tetiana Filimonchuk, Halyna Maistrenko, Vitalii Sitnikov

Abstract. Topicality. Modern people are constantly looking for new ways to use energy to improve their lives, so the demand for it is growing. In most cases, it is difficult for companies and industries to control all their devices at the same time, which can lead to a loss of electricity at any time. As a result, operating costs will be higher than necessary. In addition, the loss of power contributes to global warming through the release of carbon when energy is generated by burning coal, gas, and oil. Thus, solutions are needed to address these issues. **The purpose of this work** is to analyse existing methods for detecting anomalies in data to solve the problem of excessive electricity consumption and to warn of critical values in the indicators of electricity consumed by various devices and electrical equipment. **The object of the study** is the process of detecting atypical values or significant deviations in electricity consumption by such parameters as voltage, current strength and frequency, power. **The subject** of the study is models and methods for detecting anomalies in data. **Results.** After a thorough analysis of each of the above anomaly detection methods, new opportunities for solving the energy consumption problem open up. For example: combining several methods into one; developing a machine learning model based on one or more methods, training on test data and, in the future, processing real energy consumption data to identify atypical values, with the ability to record the date and time of anomalies, and build various graphs based on this information. **Conclusion.** The anomaly detection methods discussed here can prevent high electricity consumption to achieve energy savings, remind users to identify faulty electrical appliances or change incorrect electricity consumption patterns, reduce users' energy costs, and promote awareness of electricity safety.

Keywords: anomaly detection, energy consumption, statistical data, machine learning, energy efficiency, clustering, methods.

UDC 004.9 **doi: 10.26906/SUNZ.2024.1.050**

G. Golovko, O. Rudenko, A. Batrachenko, R. Kyzymenko

National University «Yuri Kondratyuk Poltava Polytechnic», Poltava, Ukraine

ORGANIZATION OF INFORMATION PROTECTION AT THE «DRIVE PETROL» ENTERPRISE USING A CRYPTOGRAPHIC ALGORITHM AES

Abstract. In today's digital age, information protection is becoming a dominant task, as the number of threats in the field of cyber security is constantly increasing. In this context, the implementation of effective protection means, among which encryption takes a key place, becomes especially important. The AES (Advanced Encryption Standard) cipher appears to be an exceptionally powerful tool aimed at ensuring data privacy. Information security is an extremely important aspect in the digital age, where cyber security threats are constantly increasing. In this context, encryption becomes a necessity, and the AES (Advanced Encryption Standard) cipher appears to be an exceptionally effective tool for ensuring data privacy. AES is used to protect information by converting it into cryptographically unreadable form. Due to the high degree of complexity and the possibility of using keys of different lengths, from 128 to 256 bits, AES guarantees a high level of security. Its resistance to attacks provides reliable protection against unauthorized access. One of the key advantages of AES is its versatility – it is used in a variety of industries, including finance, medicine, telecommunications, and more. The cipher has high performance, which makes it the optimal solution for protecting confidential information in a world where security becomes a priority. The application of AES not only protects data from unauthorized access, but also contributes to the overall level of security in the digital environment, ensuring excellent compatibility with various industries and user needs.

Keywords: cryptography, functions, cypher, aes, operator, algorithm.

Introduction

The problem of information protection is not new. It appeared long before the advent of computers. The rapid improvement of computer technologies also affected the principles of building information protection. From the very beginning of its development, information security systems were developed for military departments. Disclosure of such information could lead to huge casualties, including human casualties. Therefore, confidentiality (that is, non-disclosure of information) was given special attention in the first security systems. It is obvious that only their complete encryption can reliably protect messages and data from disclosure and interception. The main feature of the current situation is that the most important task today is the protection of information in computer networks [1].

The widespread introduction of computers in all types of activities, the constant increase in their computing power, the use of computer networks of various scales have led to the fact that the threat of loss of confidential information in data processing systems has become an integral part of almost any activity. The principle of modern information protection can be expressed as follows - the search for an optimal relationship between availability and security. A fully secured computer is one that is locked in an armored room in a safe, not connected to any network (not even electrical) and turned off. Such a computer has absolute protection, but it cannot be used. In this example, the requirement of availability of information is not fulfilled. The "absoluteness" of protection is hindered not only by the need to use protected data, but also by the complexity of protecting systems. [1]

Main part

1. Organization of information protection at the enterprise. An official - the head of the security department - is appointed to manage the means of information protection at the enterprise. His duties and competence:

- creation of a system of delimiting access and means of protection of the facility;
- protection of information from leakage through technical channels;
- implementation of technical measures for information protection.

2. Peculiarities of the implementation of the access demarcation system. In the access delimitation system, a dispatcher must be used, which performs access delimitation in accordance with the principle of delimitation.

Demarcation of access to information objects is carried out in accordance with the authorities of the subjects.

The basis of such demarcation is the selected access control model implemented by the access manager.

The manager ensures the implementation of the rules for demarcating the access of subjects to access objects, which are stored in the database of authorizations and access characteristics.

A request for subject access to some object is sent to the database management and event registration unit. The authority of the subject and the characteristics of the object are analyzed in the decision-making block. According to the results of the analysis, a signal of permission or refusal of permission is formed ("Allow", "Reject").

If the number of "Reject" signals exceeds a given level (for example, 5 times), which is fixed by the registration unit, then the decision-making unit signals "Unauthorized access".

Based on this signal, the security system administrator can block the subject's work to find out the cause of such violations.

Distribution of information is shown in Table 1.

Table 1 – **Distribution of information**

Types of information / Departments	General information	Personal information	Financial information	Economic information	Legal information	Technical information
CEO	+	+	+	+	+	+
Accounting	+	-	+	-	-	-
Head of HR department	+	+	-	-	-	-
Chief designer	+	-	-	-	-	+
Level of confidentiality of information	N	N	S	S	S	R

N – not secret; S – secret; R – restricted

3. Information protection methods. Such a classification of information protection methods is usually considered:

1) legislative;
2) organizational;
3) technical;
4) software;
5) moral and ethical;
6) cryptographic [2].

The initial stage of the development of computer security is strongly connected with cryptography. The main conditions of information security are its availability and integrity. In other words, the user can at any time request the set of services he needs, and the security system must guarantee its correct operation. Any file or system resource, subject to compliance with access rights, should be available to the user at any time. If some resource is unavailable, then it is useless. Another task of protection is to ensure the immutability of information during its storage or transmission. This is the so-called integrity condition.

Performing encryption and decryption procedures, in any information process system, slows down data transfer and reduces their availability, because the user will wait too long for his "reliably protected" data, which is unacceptable in some modern computer systems. Therefore, the security system must first of all guarantee the availability and integrity of information, and then (if necessary) its confidentiality.

The AES cipher is a symmetric block encryption algorithm (block size 128 bits, key 128/192/256 bits), a finalist in the AES competition and adopted as the US encryption standard by the US government. The choice fell on AES with the expectation of widespread use and active analysis of the algorithm, as was the case with its predecessor, DES.

The US National Institute of Standards and Technology published the preliminary AES specification on October 26, 2001, after five years of preparation. On May 26, 2002, AES was announced as the encryption standard.

As of 2009, AES is one of the most widely used symmetric encryption algorithms [3].

4. Access restrictions. Demarcation of access is to provide each registered user with the opportunity to freely access information within the limits of his authority and to exclude the possibility of exceeding these authority. For each user, his authority regarding files and directories is established. [4]

The Bell-LaPadulla model is an access control and management model that is based on the mandated access control model (Tabl. 2).

Table 2 – **The Bell-LaPadulla model**

Types of information / Departments	General information	Personal information	Financial information	Economic information	Legal information	Technical information
CEO	F	F	F	F	F	F
Accounting	R	N	F	N	N	N
Head of HR department	R	F	N	N	N	N
Chief designer	R	N	N	N	N	N

R – reading; N – no access; F – full access

The model analyzes the conditions under which the formation of information flows from subjects with a higher level of access to subjects with a lower level of access is impossible [4–6].

In the developed application, accounts were created for each user (Tabl. 3).

Table 3 – **Accounts for each user**

Office	Login (comp. name)	Password
State Department	deputydir	dpceo001
CEO	gendir	ceo001
Secretary	secret1	secret001
Secretary	secret2	secret002
Secretary	secret3	secret003
Chief Accountant	leadbuh2	leadbuh001

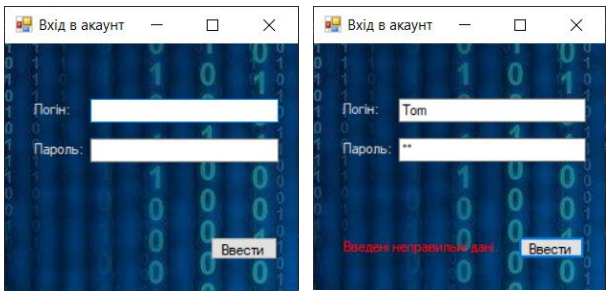
5. Presentation of the application/ After starting the program, we see the system login window (Fig. 1, a), if you do not enter the login and password, the program will not allow you to continue working (Fig. 1, b). Logging in is done using accounts for each user.

The menu window is shown in Fig. 2.

To work with the program, you need to create keys immediately (Fig. 3).

Then you can export these keys to your device for further work with them. In this case, when entering the program again, you will not have to create them again, you will only need to import them.

To encrypt a file, you need to click on "Encrypt file" (Fig. 4, 5). To decrypt the file, you need to click on "Decrypt file" (Fig. 6).



a b

Fig. 1. Login window

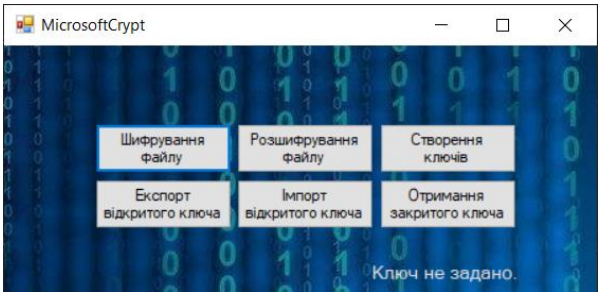


Fig. 2. Menu window

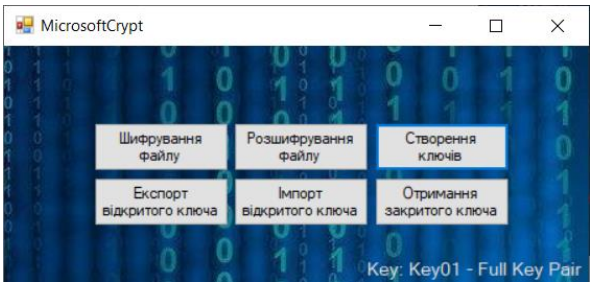


Fig. 3. An example of creating keys

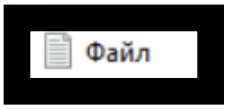


Fig. 4. File for encryption

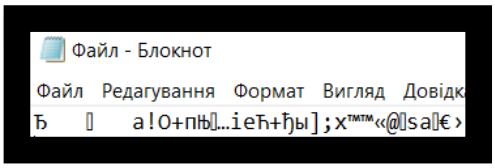


Fig. 5. File after encryption

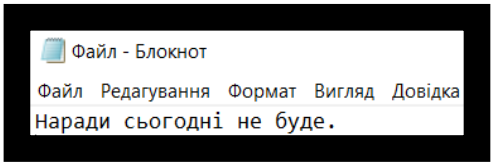


Fig. 6. File after decryption

Conclusion

As a result of the implementation of the scientific article, the project of organized means of information protection in the company "Drive Petrol" was developed. Demarcation of access was implemented, an access matrix and a mandated model of access to information were built. Information is also protected at the registry level and at the password level.

User accounts are created and an antivirus program is selected to protect the company's information from viruses, taking into account all the advantages and disadvantages.

An application has been developed, the main task of which is to encrypt data using the AES cipher.

REFERENCES

1. Concept, essence, meaning of information protection. URL: http://www.infobezpeka.com/publications/?id=102.
2. Means of information protection. URL: https://stud.com.ua/94403/informatika/zasobi_zahistu_informatsiyi.
3. AES encryption. [Electronic resource]. – Access mode: http://kriptografea.narod.ru/AES.html.
4. Devyanin P.N. Safety models of computer systems: A textbook. Akademiya, 2005. P. 55-66.
5. Control, navigation and communication systems.
6. Golovko G., Matiashenko A., Solopihin N. Data encryption using XOR cipher.

Received (Надійшла) 29.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Організація захисту інформації на підприємстві «Драйв Петрол» з використанням криптографічного алгоритму AES

Г. Головко, О. Руденко, А. Батраченко, Р. Кизименко

Анотація. В сучасному цифровому віці захист інформації стає домінуючим завданням, оскільки постійно зростає кількість загроз у сфері кібербезпеки. У цьому контексті особливо важливим стає впровадження ефективних засобів захисту, серед яких ключове місце займає шифрування. Шифратор AES (Advanced Encryption Standard) видається винятково потужним інструментом, спрямованим на забезпечення конфіденційності даних. Захист інформації – це надзвичайно важливий аспект у цифровому віці, де загрози кібербезпеки постійно зростають. У цьому контексті шифрування стає необхідністю, а шифратор AES (Advanced Encryption Standard) видається винятково ефективним інструментом для забезпечення конфіденційності даних. AES використовується для захисту інформації шляхом перетворення її у криптографічно нерозбірливий вигляд. Завдяки високій ступені складності та можливості використання ключів різної довжини, від 128 до 256 біт, AES гарантує високий рівень безпеки. Його стійкість до атак забезпечує надійний захист від несанкціонованого доступу. Однією з ключових переваг AES є його універсальність – він застосовується в різних галузях, включаючи фінанси, медицину, телекомунікації тощо. Шифратор володіє високою продуктивністю, що робить його оптимальним рішенням для захисту конфіденційної інформації у світі, де безпека стає пріоритетом. Застосування AES не лише забезпечує захист даних від несанкціонованого доступу, але й сприяє підвищенню загального рівня безпеки в цифровому середовищі, забезпечуючи відмінну сумісність із різними індустріальними галузями та потребами користувачів.

Ключові слова: криптографія, функції, шифр, aes, оператор, алгоритм.

УДК 004.7 **doi: 10.26906/SUNZ.2024.1.053**

В. О. Горбачов, О. А. Янковський, В. Р. Діян, Д. І. Балінський

Харківський національний університет радіоелектроніки, Харків, Україна

МЕТОДИ ПРОЕКТУВАННЯ СИСТЕМИ ДОКУМЕНТООБІГУ УНІВЕРСИТЕТУ

Анотація. Через постійне збільшення обсягів даних, які зберігаються та обробляються інформаційними системами, потрібне ретельне проектування архітектури систем, щоб уникнути зниження гнучкості та якості обробки даних. Метою даної роботи є проведення аналізу технологій розробки архітектури та програмного забезпечення паралельного розподіленого застосунку. Пропонується розробка та впровадження електронної системи документообігу в адміністративній системі вищого навчального закладу. У статті наведено аналітичне моделювання методом COMET системи документообігу вищого навчального закладу. Наукова новизна цього дослідження полягає в тому, що отримана модель дозволяє провести подальше математичне моделювання системи з метою оцінки параметрів реальної системи та визначення вузьких місць. Така модель представляє практичну цінність для подальших досліджень, розширення функціональності системи, а також для навчання нових співробітників вищого навчального закладу.

Ключові слова: паралельні застосунки, розподілені застосунки, архітектурне проектування, моделювання паралельних систем.

Вступ

Концепція паралельних процесів, є основою проектування паралельних застосунків. Паралельний застосунок складається з багатьох завдань, що виконуються одночасно. Концепції проектування паралельних завдань можна застосувати також до розподілених застосунків. Основна складність полягає в тому, щоб розбити застосунок на паралельно виконувані завдання та надати засоби обміну повідомленнями та синхронізації цих завдань між собою.

Для розподілених застосунків потрібно подбати і про інші характеристики. Розподілений застосунок – це паралельний застосунок, який виконується у розподіленому середовищі, що складається з кількох географічно рознесених вузлів. Розподілений застосунок складається з паралельних процесів, що працюють у різних вузлах. У той самий час, кожен процес може мати кілька потоків, виконуваних у тому вузлу. Розподілена обробка має такі переваги:

- гнучка структура - один і той же застосунок може мати різні структури, при розміщенні його на відповідному числі вузлів;

– більш локалізоване управління та адміністрування – розподілену підсистему, що виконується на своєму власному вузлу, можна спроектувати так, що вона буде автономною, тобто практично незалежною від інших підсистем, що працюють на інших вузлах;

- поступове розширення системи - якщо навантаження сильно зростає, систему легко розширити за рахунок додавання нових вузлів;

– зменшення витрат – найчастіше розподілене рішення виявляється дешевшим від централізованого, особливо якщо взяти до уваги стрімко зростаючу продуктивність мікрокомп'ютерів;

– балансування навантаження – в деяких додатках загальне навантаження на систему може бути розподілене між різними вузлами;

– зменшення часу відгуку – запити користувачів локальних систем обробляються швидше.

Об'єктом дослідження у роботі є адміністративна розподілена система. Системи такого роду задаються переліком підрозділів, їх призначенням, функціями та зв'язками між собою. Одним з найважливіших чинників у тому функціонуванні є документи. У складній системі щорічно циркулюють десятки різних типів офіційних документів, які керують, супроводжують, інформують кожен із підрозділів системи. Для наочності як приклад адміністративної розподіленої системи обрано систему вищого навчального закладу.

Мета статті – аналіз інформаційної системи вищого навчального закладу та розробка моделі електронної системи документообігу. У зв'язку з постійним розвитком та укрупненням перспективних систем, нині на супровід документообігу витрачається дедалі більший людський ресурс. Отже, проблема його оптимізації у складних адміністративних системах набуває все більшої актуальності. З розвитком інформаційних технологій з'явилася можливість більш ефективної організації процесу документообігу із застосуванням баз даних, засобів електронного підпису та інших програмних засобів. Створення оптимальної електронної системи документообігу передбачає попередній аналіз предметної галузі та моделювання її аспектів.

Для досягнення поставленої мети вирішено такі завдання. Досліджено провідні методи системного аналізу та серед них обрано найбільш підходящий метод для моделювання розподіленої адміністративної системи.

На підставі проведеного аналізу обраним методом побудовано модель системи документообігу у вищому навчальному закладі.

Таку систему зручно уявити як систему з розподіленими модулями, також виділивши клієнтські та серверні частини. Для розробки моделі системи документообігу серед великої кількості різноманітних методів системного аналізу та проектування вибрано метод COMET [1]. Визначальними чинниками вибору цього є його орієнтованість на розробку систем реального часу і розподілених систем. Практична цінність роботи полягає у отриманні моделі структури адміністративної системи для подальшого аналізу та застосування в відповідних системах.

Аналіз сучасних методів системного аналізу проектування складних систем

Стандарт ISO/IEC 12207 визначає структуру життєвого циклу програмної системи, що містить процеси, дії та завдання, які мають бути виконані під час створення інформаційної системи. Кожен із процесів характеризується певними завданнями та методами їх вирішення, вихідними даними, отриманими на попередньому етапі, та результатами. У результаті еволюції нові методи проектування продовжують з'являтися на вирішення завдань, не вирішуваних попереднім поколінням методів.

Центральне місце у різноманітті методів системного аналізу та проектування займають методи структурованого проектування. В результаті застосування структурних методів до паралельних систем і систем реального часу з'явилися різні методи системного аналізу, еволюціонуючи згодом об'єктно-орієнтовані методи. Модель водоспаду (або класична каскадна модель) стала першою і найбільш використовуваною моделлю, яка структурує процес розробки. В основі цього методу лежить поетапний підхід до розробки систем при припущенні, що кожен попередній етап завершується на початок наступного, і не відбувається повернення на попередні кроки проектування. При використанні такої моделі проблемою виявилася можливість невідповідності кінцевого продукту та початкових вимог, які до нього пред'являлися.

Для виключення подібних помилок розроблено модифікацію моделі водоспаду. Так, метод тимчасових прототипів використовується під час створення складних інтерфейсів для інтерактивних інформаційних систем [1]. Користувач, працюючи з макетом, з кожною ітерацією більш точно формулює свої вимоги до кінцевого продукту. Однак недоліком методу є підвищення витрат через проектування прототипів. У свою чергу, в рамках інкрементного методу, завдання розбивається на відносно незалежні складові, що проектуються окремо. Проміжні прототипи створюються послідовно, починаючи з ранніх стадій, з подальшим розширенням функціональності до отримання готової системи. Але в цьому випадку розподіл на модулі уповільнює та ускладнює процес проектування, тому що необхідно забезпечити їх взаємодії між собою та керувати отриманою складною системою.

Для вирішення вищезазначених проблем запропоновано спіральну модель, в якій кожен виток спіралі відповідає створенню фрагмента або нової версії програмного забезпечення. Розробка ітераціями відбиває об'єктивно існуючий спіральний цикл створення систем. Невиконана на якомусь етапі робота буде виконана на наступній ітерації. Спіральна модель легко управляє процесом розробки при змінах вимог до проекту, змін параметрів проекту або тимчасових затримках. Ідея того, що процес роботи над проектом може складатися з циклів, що проходять одні й самі етапи, стала основою більшості сучасних моделей. В даний час набула широкого поширення об'єктно-орієнтована методологія. Метод OOSE (Object-Oriented Software Engineering) застосовує об'єктно-орієнтований аналіз, проектування та програмування [2]. На основі цього методу створено метод RUP (Rational Unified Approach) – один із найпоширеніших методів комплексного управління процесом розробки, який систематизує процес створення програмного забезпечення із застосуванням UML (Unified Modelling Language) [3].

Розробка моделі документообігу вищого навчального закладу за допомогою методу COMET

Розробка моделі документообігу вищого навчального закладу за допомогою COMET заснована на ітеративному процесі проектування та концепції прецедентів [1]. Вибір методу COMET як метод системного аналізу для розробки моделі документообігу обґрунтований його адаптацією для розподілених додатків та систем реального часу.

Відповідно до правил методу, проектування будь-якої системи починається з моделювання вимог. Система розглядається як чорна скринька, і враховуються лише її зовнішні характеристики. Для демонстрації першого етапу моделювання на рисунку 1 представлений прецедент переведення студентів на наступний курс, що є складовою прецеденту контролю успішності студентів.

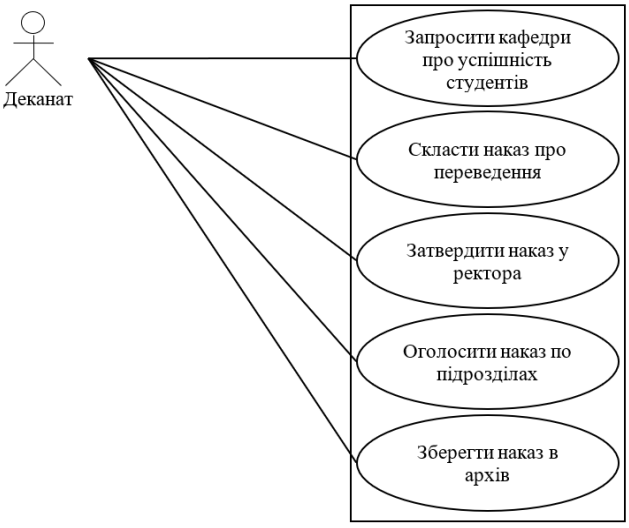


Рис. 1. Прецедент «Перевести студентів на наступний курс»

У термінах акторів та прецедентів визначено функціональні вимоги до системи. Кожен прецедент визначає послідовність взаємодій між кількома акторами.

Для опису варіантів використання системи потрібно було описати групи прецедентів про здобуття вищої освіти, про порядок навчання студентів викладачами, про контроль навчального процесу деканатом і взаємодіями, що призводять до необхідного результату, та іншими найбільш можливими подіями, що призводять до альтернативних результатів. Прецеденти, вказані для кожного актора у системі, визначають всі вимоги цього актора до системи. Так визначається потрібна функціональність системи без розкриття внутрішньої структури.

Предметна область розглянута на етапі аналітичного моделювання, внаслідок чого побудовано статичну та динамічну моделі системи. За допомогою статичної моделі задаються структурні відносини між класами предметної галузі щодо прецедентів. Для опису такої моделі використовуються діаграми класів, діаграми станів, діаграми кооперації та послідовності. Модель складена з об'єктів (за критеріями розбиття на об'єкти) та їх взаємодій. В результаті отримано структурне подання інформаційних аспектів системи з класами, їх атрибутами та відносинами між ними. На рис. 2 показано статичну модель системи вищого навчального закладу із відносинами між інформаційно насиченими класами, призначеними для зберігання даних. Оскільки в цій системі використовується багато різноманітних даних, основну увагу при моделюванні приділено сутнісним класам. На рис. 2 показані класи для найбільш значущих сутностей, таких як "Розклад", "Навчальний план", "Наказ" та "Відомість".

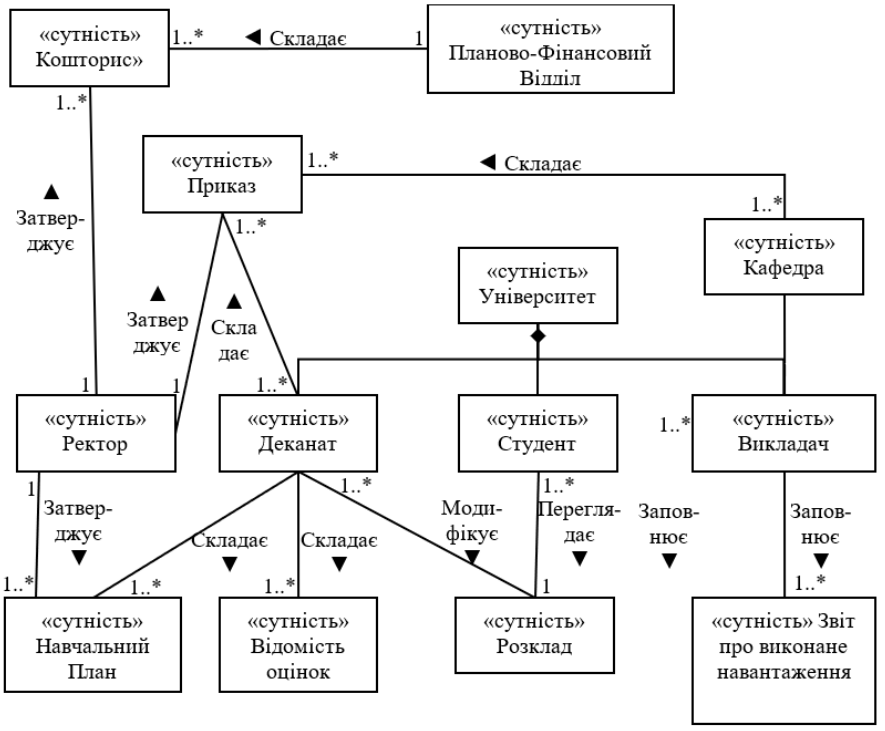


Рис. 2. Концептуальна статична модель ЕСД з власними класами

Модель контексту системи уточнена за допомогою діаграм класів, на яких відображаються інтерфейси між системою та зовнішніми класами. Такі діаграми потрібні для проектування систем, що управляють зовнішніми пристроями вводу-виводу та зовнішніми системами. На рис. 3 наведено статичну модель досліджуваної предметної області з діаграмами класів контексту електронної системи документообігу.

За результатами статичного моделювання проведено декомпозицію предметної області на програмні об'єкти. Виявлені на попередніх етапах зовнішні та сутнісні класи в системі документообігу, використовуються для категоризації класів та об'єктів, які будуть реалізовані в програмній системі, на сутнісні, інтерфейсні, керуючі та пов'язані з прикладною логікою.

На рис. 4 зображено класифікацію класів системи у стереотипах «застосунок», «інтерфейс», «сутність», «прикладна логіка». За стереотип «прикладна логіка» прийнято класи алгоритмів складання розкладу, розподілу навантаження викладачів, розрахунку кошторису та ін. Сутностями є бази даних та документи, що циркулюють за системою вищого навчального закладу.

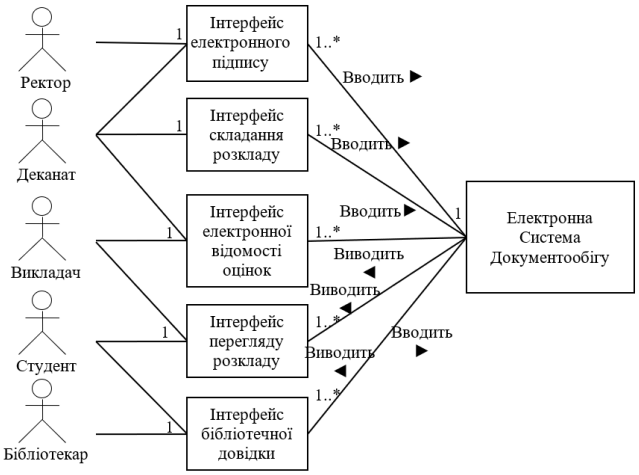


Рис. 3. Діаграма класів контексту ЕСД

Як інтерфейс вибрано стандартний інтерфейс персонального комп'ютера або терміналу. Усі нові класи додані до словника класів проектованої системи, створеного на етапі статичного моделювання.

В процесі моделювання, система вищого навчального закладу була розділена на окремі підсистеми, як показано на рис. 5.

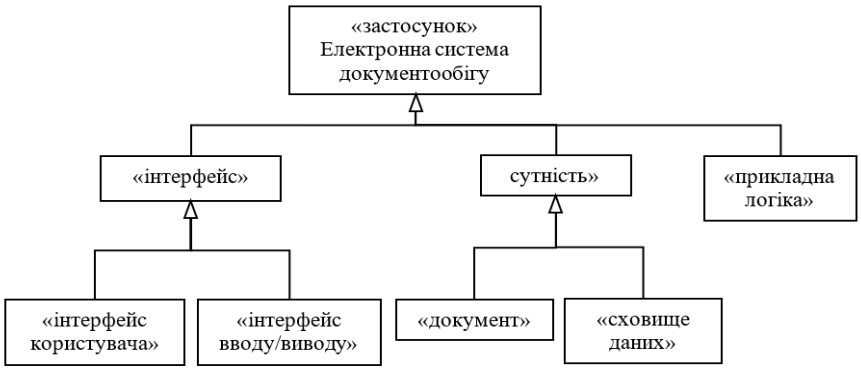


Рис. 4. Стереотипи класів ЕСД

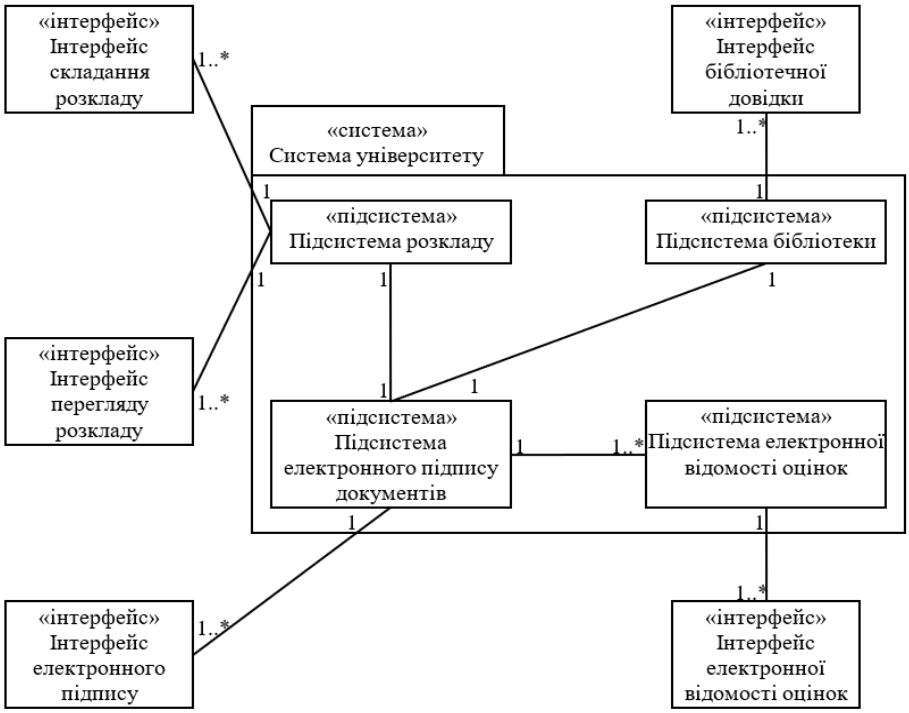


Рис. 5. Основні підсистеми та їхні зв'язки в системі університету

Підсистема розкладу включає класи та об'єкти, спільна основна функція яких полягає в попередньому зборі інформації, складанні та видачі навчального розкладу. Підсистема електронної відомості складається із засобів збору, зберігання та обробки інформації про результати екзаменаційних сесій. Підсистема електронного підпису документів має спеціальні кошти на затвердження наказів, навчальних планів, кошторисів та інших офіційних документів.

У досліджуваній системі виділено клієнтські та серверні частини. Клієнтом є частина системи, яка посилає запити до серверної частини. Сервер відповідно обробляє запит та повертає результат. Аналіз прецедентів показав, що в системі клієнтом у більшості випадків є деканат. Деканат надсилає запити на кафедру про успішність студентів та виконання викладацького навантаження, на початку кожного семестру заявку до центру складання розкладу, до бібліотеки – про наявність друкованих матеріалів. Відповідно, об'єкти з іншого боку взаємодії є серверами. Існує багато систем, функціонування яких визначається як вхідними даними, так і станами системи на попередніх кроках. При проектуванні такої системи необхідно проводити моделювання можливих її станів. У описі системи методом COMET динамічні аспекти системи відбиваються діаграмами станів. На етапі моделювання станів система має вигляд кінцевого автомата. Моделювання станів системи документообігу проведено для кожного з прецедентів за допомогою ієрархічних діаграм станів, де задано аспекти системи, що залежать від стану. Наприклад, на рис. 6 представлена діаграма станів для прецеденту «Перевести студентів на наступний курс» з боку підсистеми видання наказів.

На заключному етапі аналітичного моделювання виконано динамічне моделювання. Усі прецеденти перевіряються з метою виявлення взаємодій між об'єктами, що беруть участь. Для кожного прецеденту визначено об'єкти, що беруть участь у ньому, і розроблені діаграми взаємодій об'єктів. На підставі опису прецеденту позначається відповідними номерами порядок повідомлень. На рис. 7 зовнішня подія «Наказ затверджений» має порядковий номер 4.

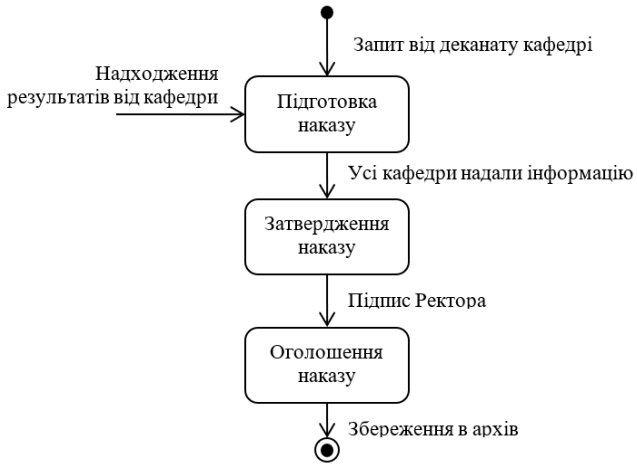


Рис. 6. Діаграма станів для підсистеми видання наказів

Рис 7. Діаграма кооперації для підсистеми видання наказів за прецедентом «Перевести студентів на наступний курс»

Об'єкт «Інтерфейс електронного підпису документів» у відповідь на це повідомлення видає 3 паралельні повідомлення «Роздрукувати документ», «Зберегти документ» та «Розіслати всім». Ці повідомлення з'являються паралельно і відповідно мають назви 4.1.a, 4.1.b і 4.1.c*. Повідомлення «Розіслати всім» поділяється на стільки копій, скільки відділів потрібно повідомити про видання нового наказу. Як показано у роботі, в результаті аналітичного моделювання системи документообігу у вищому навчальному закладі, отримано модель, що складається з низки діаграм, що описують структуру та поведінку системи. На основі виявлених прецедентів побудовано діаграму концептуальної статичної моделі документообігу, моделі сутнісних класів та контекстів для визначення інтерфейсів до системи. З отриманих діаграм виділено класи та об'єкти за стереотипами. Система розділена на підсистеми із зазначенням кількісних зв'язків між ними. Модулі, функціонування яких залежить від стану, показані на діаграмах станів. Для прецедентів на діаграмах кооперації зазначено взаємодію об'єктів у часі відповідно порядку надходження подій. Отже, в отриманих моделях відображені всі аспекти, які необхідно дослідити та врахувати під час проектування системи документообігу.

Висновки

У статті наведено аналітичне моделювання методом COMET системи документообігу вищого навчального закладу. Наукова новизна цього дослідження полягає в тому, що отримана модель дозволяє провести подальше математичне моделювання системи з метою оцінки параметрів реальної системи та визначення вузьких місць. Така модель представляє практичну цінність для подальших досліджень, розширення функціональності системи, а також для навчання нових співробітників вищого навчального закладу.

Майбутні дослідження: розробка математичних моделей для оцінка продуктивності системи документообігу; впровадження системи документообігу на розподіленій обчислювальній платформі.

СПИСОК ЛІТЕРАТУРИ

1. H. Gomaa, Designing Concurrent, Distributed, and Real-time Applications with UML, Addison-Wesley, 2000
2. Jacobson I. Object-Oriented Software Engineering. S.1.: ASM press., 1992. – 528 p.
3. Selic B., Gullekson G., Ward P. T. Real-Time Object Oriented Modelling. – S.1.: John Wiley & Sons, 1994. – 525 p.

Received (Надійшла) 29.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Design Approaches of Document Management System of University

V. Gorbachov, O. Yankovsky, V. Diyan, D. Balinskiy

Abstract. Due to the constant increase in the amount of data stored and processed by information systems, a careful design of the system architecture is required to avoid a decrease in the flexibility and quality of data processing. The purpose of this work is to analyze the architecture and software development technologies of a parallel distributed application. It is proposed to develop and implement an electronic document management system in the administrative system of a higher educational institution. The article provides analytical modeling by the COMET method of the document management system of a higher educational institution. The scientific novelty of this study is that the resulting model allows for further mathematical modeling of the system in order to estimate the parameters of the real system and identify bottlenecks. Such a model is of practical value for further research, expanding the functionality of the system, as well as for training new employees of a higher education institution.

Keywords: parallel applications, distributed applications, architectural design, simulation of parallel systems.

УДК 004.7 **doi: 10.26906/SUNZ.2024.1.058**

В. Д. Дюльгер, А. Р. Сорокін

Харківський національний університет радіоелектроніки, Харків, Україна

АНАЛІЗ МЕТОДІВ ІНТЕГРАЦІЇ ТА УЗГОДЖЕННЯ МІКРОСЕРВІСІВ В ХМАРНІЙ АРХІТЕКТУРІ

Анотація. Метою даної роботи є проведення аналізу сучасних методів інтеграції мікросервісів в хмарній архітектурі. Розділи роботи включають вивчення сучасного ландшафту хмарних архітектур та мікросервісів, їхню сутність та переваги в контексті хмарних середовищ. Детальний аналіз викликів та особливостей інтеграції мікросервісів включає розділи про контейнеризацію та оркестрацію, API-взаємодію та гібридні рішення, а також мікросервісну шину. Робота визначає роль та функціонал кожного методу, розкриває їхні переваги та виклики, що допомагає розуміти оптимальні стратегії інтеграції для розробників та організацій у сучасному інформаційному просторі.

Ключові слова: хмарні технології, мікросервісна архітектура, контейнеризація, serverless.

Вступ

В сучасному інформаційному суспільстві, характеризованому стрімким розвитком технологій, хмарні архітектури та мікросервіси визначають новий стандарт в галузі розробки програмного забезпечення. Розповсюдження хмарних рішень та використання мікросервісної архітектури відкривають безліч можливостей для оптимізації та розширення функціональності програмних продуктів. Зокрема, важливим аспектом стає аналіз методів інтеграції та узгодження мікросервісів в хмарній архітектурі. Спрямовані на взаємодію компоненти системи вимагають особливої уваги у зв'язку з різноманітністю сервісів, їхньою динамічністю та необхідністю забезпечення надійності та ефективності функціонування. Дослідження вказаної теми відкриває можливість вдосконалення стратегій розгортання мікросервісів у хмарних середовищах та сприяє оптимізації їхньої взаємодії для досягнення максимальної ефективності в різноманітних інформаційних проектах.

Мета статті – проведення аналізу методів інтеграції та узгодження мікросервісів в хмарній архітектурі та огляд перспектив. Дослідження спрямоване на виявлення оптимальних стратегій розгортання мікросервісів, їхньої ефективної взаємодії та забезпечення надійності в хмарних середовищах. Результати дослідження можуть внести вагому вклад у покращення розробки та управління програмним забезпеченням в контексті сучасних технологічних тенденцій.

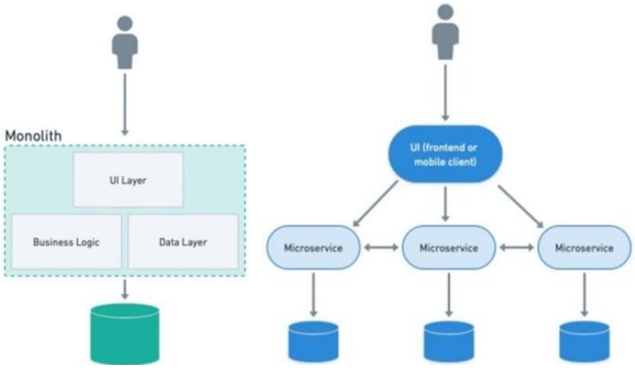
Результати досліджень

1. Сучасний ландшафт хмарних архітектур та мікросервісів. Хмарні архітектури базуються на принципах, таких як віртуалізація ресурсів, самообслуговування, широкий доступ та масштабованість. Вони надають користувачам можливість використовувати ресурси за вимогою, оптимізуючи використання обчислювальних потужностей та забезпечуючи гнучкість [2].

Публічні хмарні платформи, такі як AWS, Azure та Google Cloud, пропонують широкий функціонал для розгортання та управління ресурсами. Приватні хмари, в свою чергу, надають більший контроль та безпеку. Аналіз переваг кожного типу платформи допомагає підібрати оптимальний підхід для конкретного проекту.

Хмарні сервіси стають важливою складовою при створенні та розгортанні програмного забезпечення. Вони дозволяють зменшити час розробки, спростити процеси тестування та забезпечити високу доступність додатків. В свою чергу, мікросервісна архітектура розглядає додаток як сукупність невеликих, незалежних сервісів. Її переваги включають легку масштабованість, незалежність компонентів та покращену ефективність розробки.

Порівняння мікросервісів з монолітними рішеннями (рис. 1) вказує на переваги, такі як легка масштабованість окремих компонентів та спрощення розвитку, що робить мікросервіси більш гнучкими та підходящими для сучасних вимог ринку.



а б

Рис. 1. Монолітна (а) та мікросервісна (б) архітектура розгортання програмного забезпечення

Мікросервіси дозволяють декомпозицію системи на невеликі фрагменти, що полегшує розвиток та управління окремими компонентами. Це призводить до збільшення гнучкості та масштабованості всієї системи [4].

2. Сутність мікросервісної архітектури полягає в її гнучкості – здатності змінювати та адаптуватися до нових умов та вимог. Кожен мікросервіс може бути розгляданий як окрема одиниця, що може бути легко масштабована горизонтально для забезпечення оптимальної продуктивності. Мікросервісна архітектура надає високий рівень незалежності від конкретних хмарних інфраструктур. Це дозволяє

ефективно використовувати можливості різних хмарних сервісів без значних модифікацій коду мікросервісів [5]. В хмарних середовищах мікросервіси можуть використовувати еластичність для автоматичного масштабування ресурсів в залежності від завдань та навантаження. Це сприяє оптимізації використання ресурсів та підвищує ефективність систем. Хмарні платформи надають інструменти для керування та моніторингу мікросервісами, що спрощує розгортання, керування життєвим циклом та підтримку цієї архітектури. Це робить хмарні середовища ідеальними для впровадження мікросервісної архітектури.

3. Виклики та особливості інтеграції мікросервісів в хмарній архітектурі В умовах хмарної архітектури важливо вирішити завдання забезпечення ефективної та безперебійної комунікації між розподіленими мікросервісами. Перехід до хмари може викликати проблеми зі швидкістю та надійністю мережевого обміну.

Однією з ключових особливостей є впровадження ефективних засобів безпеки для захисту мікросервісів в хмарних середовищах. Це включає управління доступом, шифрування та моніторинг безпеки [1]. Необхідно узгоджувати інтерфейси та протоколи між мікросервісами для ефективної взаємодії в хмарних умовах, де різноманітні технології можуть викликати проблеми сумісності.

Забезпечення ефективного моніторингу та управління ресурсами мікросервісів в хмарі важливо для забезпечення стабільності та високої доступності системи [1]. Інтеграція мікросервісів в хмарній архітектурі вимагає розробки гнучких та відмовостійких стратегій, щоб ефективно реагувати на зміни в середовищі та вирішувати можливі проблеми.

4. Аналіз сучасних методів інтеграції мікросервісів в хмарній архітектурі Контейнеризація та оркестрація стали ключовими компонентами для інтеграції мікросервісів у хмарних середовищах [3].

Використання контейнерів, таких як Docker, дозволяє ізолювати мікросервіси та їх залежності, забезпечуючи консистентність середовищ тестування та розгортання. Інструменти оркестрації, такі як Kubernetes, надають можливість автоматизованого керування та масштабування контейнеризованих мікросервісів. Це полегшує управління та підтримку великої кількості сервісів у хмарних обчислювальних середовищах.

Переваги:

- Ізоляція та портативність — контейнеризація дозволяє ізолювати мікросервіси, забезпечуючи консистентність середовища в різних стадіях розробки;

- Спрощення розгортання та масштабування — ркестраційні інструменти, такі як Kubernetes, автоматизують процеси розгортання та масштабування, що покращує ефективність управління.

Недоліки:

- Складність конфігурації — налаштування контейнерів та їхніх параметрів може виявитися складним завданням, особливо для великих систем;

- Велика кількість компонентів — використання оркестраторів може призвести до значного збільшення кількості компонентів у системі, що може збільшити складність управління.

API-взаємодія є важливим аспектом для успішної інтеграції мікросервісів в хмарній архітектурі.

Стандартизовані та добре задокументовані API дозволяють ефективно обмінюватися даними та взаємодіяти між мікросервісами. Використання REST або GraphQL дозволяє підтримувати гнучкі та ефективні точки доступу [3]. Інтеграція гібридних рішень дозволяє комбінувати використання власних та зовнішніх мікросервісів. Це дає можливість підтримувати специфічні функції власного бізнесу та одночасно використовувати зовнішні сервіси для покращення функціональності.

Переваги:

- Спрощена взаємодія — Використання API дозволяє просто та ефективно забезпечити взаємодію між мікросервісами, знижуючи залежності;

- Гнучкість — Гібридні рішення дозволяють комбінувати внутрішні та хмарні ресурси, забезпечуючи гнучкість і резервування.

Недоліки:

- Проблеми безпеки — забезпечення безпеки при обміні даними через API може бути складною задачею, особливо при великій кількості зовнішніх взаємодій;

- Можливість виникнення залежностей — гібридні рішення можуть призводити до виникнення складних залежностей між внутрішніми та зовнішніми компонентами.

5. Мікросервісна шина (Microservices Bus). Мікросервісна шина є іншим ефективним методом інтеграції мікросервісів у хмарних архітектурах. Шина дозволяє забезпечити взаємодію між мікросервісами, служить посередником для обміну повідомленнями та забезпечує асинхронну комунікацію між компонентами системи [4]. Застосування шини полегшує розширення функціональності та додає гнучкості у систему, а також спрощує впровадження змін у окремих мікросервісах. Аналіз та вибір оптимального методу інтеграції залежить від конкретних потреб проекту, розміру системи та вимог до її ефективності та масштабованості.

Переваги:

- Централізоване керування — мікросервісна шина дозволяє централізовано керувати взаємодією мікросервісів та зменшити зв'язок між ними;

- Спрощення розширення — додавання нових сервісів може бути спрощеним завданням, оскільки вони можуть взаємодіяти через шину.

Недоліки:

- Однопоточність — використання централізованої шини може призвести до однопоточності, де збільшення навантаження може стати обмежуючим фактором;

- Потенційний одиничний пункт відмови — централізована шина може стати одиничним пунктом відмови, якщо не враховані відповідні механізми для врегулювання цього ризику. Цей аналіз допомагає визначити оптимальний метод інтеграції мікросервісів, враховуючи конкретні потреби та обмеження конкретного проекту чи організації.

6. Узгодження мікросервісів у хмарному середовищі. У хмарній архітектурі, де мікросервіси можуть бути розташовані в різних місцях, забезпечення синхронізації даних стає критичним завданням. Розгляд методів для обміну та оновлення даних між мікросервісами, таких як подійно-орієнтована архітектура чи використання асинхронних повідомлень, є важливим для забезпечення консистентності [6]. Аналіз концепції CAP (Consistency, Availability, Partition tolerance, рис. 2) допомагає визначити, як досягти балансу між доступністю, консистентністю та стійкістю при роботі з розподіленими даними. Використання методів реплікації та версіонування даних може покращити консистентність в мікросервісній архітектурі [6].

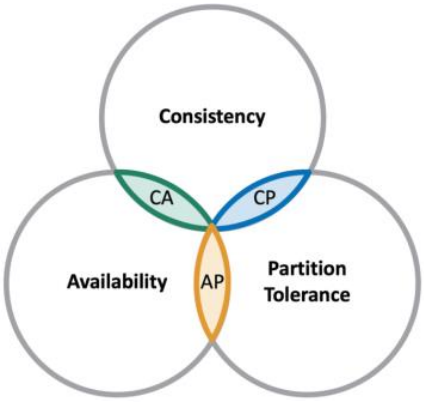


Рис. 2. Концепція Consistency, Availability, Partition tolerance

7. Перспективи розвитку та майбутні тренди. Хмарна архітектура та мікросервіси є динамічним полем розвитку, і майбутні тренди обіцяють захоплюючі перспективи для подальшого вдосконалення інтеграції мікросервісів у хмарі. Деякі ключові аспекти, які можуть визначати майбутній шлях розвитку, включають:

- розширення обчислювальних можливостей на "краї" мережі (Edge) та в туманних обчисленнях (Fog) стає важливим трендом. Це дозволяє обробку даних ближче до їх джерела, покращуючи продуктивність та знижуючи затримки [7];

- розвиток технологій контейнеризації, таких як Kubernetes, та Serverless архітектур дозволяє ефективно використовувати ресурси, спрощує розгортання та забезпечує гнучкість роботи мікросервісів;

- тренд до більшої декомпозиції на рівні функціональності, коли мікросервіси стають ще меншими та більш спеціалізованими, допомагає забезпечити гнучкість, але вимагає удосконаленої системи управління та моніторингу;

- інтеграція штучного інтелекту та машинного навчання в мікросервісні архітектури дозволяє створювати інтелектуальні та автономні системи, які можуть адаптуватися до змін у реальному часі [5];

- підвищення безпеки включає в себе вдосконалення методів автентифікації, контролю доступу та захисту даних у розподілених середовищах, щоб запобігти загрозам кібербезпеки.

Майбутнє інтеграції мікросервісів у хмарній архітектурі обіцяє новаторські рішення та розвиток технологій для підтримки високоефективних та гнучких систем. Вивчення цих тенденцій може допомогти готувати архітекторів та розробників до майбутніх викликів і можливостей у цьому еволюційному області.

Висновки

В статті розглянуто використання мікросервісної архітектури в поєднанні з хмарними технологіями, що відкриває нові горизонти для розвитку програмного забезпечення та обслуговування сучасних бізнес-процесів. Ці підходи дозволяють підприємствам бути гнучкими, швидкореагуючими та конкурентоспроможними в умовах постійних змін технологічного середовища та ринкових вимог.

СПИСОК ЛІТЕРАТУРИ

1. Newman, S. (2015). "Building Microservices: Designing Fine-Grained Systems." O'Reilly Media.
2. Lewis, J., & Fowler, M. (2014). "Microservices: a definition of this new architectural term." https://martinfowler.com/articles/microservices.html
3. Richardson, C. (2018). "Microservices Patterns: With Examples in Java." Manning Publications.
4. Wiggins, A. (2016). "Microservices in Action." Manning Publications.
5. Dragoni, N. et al. (2017). "Microservices: yesterday, today, and tomorrow." Communications of the ACM, 60(6), 85-93.
6. Armbrust, M., Stoica, I., Zaharia, M., Fox, A., Griffith, R., Joseph, A. D., ... & Zaharia, M. (2010). "A view of cloud computing." Communications of the ACM, 53(4), 50-58.
7. Dikaiakos, M. D., Katsaros, D., Mehra, P., Pallis, G., & Vakali, A. (2009). "Cloud computing: Distributed internet computing for IT and scientific research." Journal of Internet Services and Applications, 1(1), 7-18.

Received (Надійшла) 22.11.2023
Accepted for publication (Прийнята до друку) 24.01.2024

Analysis of methods for integrating and coordinating microservices in cloud architecture

V. Diulher, A. Sorokin

Abstract. The purpose of this work is to analyze modern methods of integrating microservices into cloud architectures. Sections of the paper include the study of the modern landscape of cloud architectures and microservices, their essence and advantages in the context of cloud environments. A detailed analysis of the challenges and features of microservices integration includes sections on containerization and orchestration, API interaction and hybrid solutions, and the microservice bus. The work defines the role and functionality of each method, reveals their advantages and challenges, which helps to understand the optimal integration strategies for developers and organizations in the modern information space.

Keywords: cloud technologies, microservice architecture, containerization, serverless.

УДК 004.9 **doi: 10.26906/SUNZ.2024.1.061**

Д. О. Дяченко, А. С. Гук, О. П. Міхаль

Харківський національний університет радіоелектроніки, Харків

МОДЕЛЮВАННЯ РЕСУРСОВІДНОВЛЕННЯ РОЗПОДІЛЕНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

Анотація. Актуальність. В роботі проаналізована проблематика, яка пов'язана з моделюванням поведінки і взаємодії об'єктів живої природи в біологічних системах. Відзначено актуальність розробок в даному напрямі моделювання, який є стиковим між біологією та загальною теорією систем. Конкретизовано коло питань, що підлягають моделюванню в рамках цієї роботи. Клітка живого організму має ресурс по числу поділок. Даний ресурс витрачається в процесі життєдіяльності організму. Цікавим є вивчення засобами комп'ютерного моделювання обмежень життєдіяльності багатоклітинного організму, пов'язаних з витрачанням даного ресурсу. Комп'ютерне моделювання дозволяє істотно економити ресурси, що витрачаються в процесі пізнання і освоєння природи, зокрема в ході розробки прикладних технічних та інформаційних систем. Клітинні автомати дозволяють абстрагуватися від багатьох несуттєвих рис об'єктів, зосереджуючи переважно увагу на циклічності процесів, просторовій розподіленості об'єктів, дискретності взаємної незалежності подій. В ресурсних клітинних автоматах додатково береться до уваги тимчасова обмеженість процесів: скорочення можливостей для подальшого їх продовження. Сукупність зазначених якостей створює формальну причинно-наслідкову середу, зручну для розгляду досліджуваних об'єктів. Ресурс - є значення певного параметра (або декількох параметрів), критичних для продовження функціонування певного елемента клітинного автомата. Ресурс може бути заданий спочатку і може спадати в процесі функціонування клітинного автомата. Сформульовано вимоги до розроблюваної моделі, які обумовлені характером, обсягом і заходом ілюстративності інформації, яка передбачається до отримання. Мова виконання моделі - Python. Алгоритм моделі представлений у вигляді покрокового опису, з необхідним коментуванням, використовуваним далі при розробці програмної реалізації моделі. Представлена і прокоментована структурна схема моделі. Показано відповідність блоків структурної схеми покрокового опису алгоритму моделі. **Мета роботи** – розробка і реалізація клітинно-автоматної комп'ютерної моделі для дослідження поведінки структури в умовах знакозмінної зміни параметрів, що визначають ресурсність.

Ключові слова: клітинний автомат, ресурс, розподілена інформаційна система, ресурсовідновлення, протокол.

Вступ

З кожним роком обчислювальна можливості людини збульшуються. Існує багато способів її використання. Один з них - моделювання все більш складних процесів [1]. Наразі в цьому напрямку розвиваються різні підходи, і одним з них є теорія клітинних автоматів. Клітинний автомат (КА) - це дискретна динамічна система, поведінка якої повністю визначається в термінах локальної взаємозалежності станів такої системи. Простір представлений рівномірною сіткою, кожна комірка якої містить кілька бітів даних.

Правила еволюції представлені набором правил, де на кожному кроці кожна клітинка обчислює новий стан зі станів сусідніх клітинок. При правильному наборі правил такого простого механізму роботи достатньо для підтримки всіх ієрархічних структур і явищ. Клітинні автомати є корисною моделлю для багатьох досліджень у природничих науках. Подібно до машин Тюрінга для послідовних обчислень, вони формують загальну парадигму для паралельних обчислень.

Середовище, представлене клітинними автоматами, має великий потенціал у моделюванні кластерів взаємопов'язаних однорідних об'єктів. Це моделювання фізичних процесів у фізиці елементарних частинок та ядерній фізиці, моделювання потоків рідини, моделювання взаємодіючих клітинних систем у біології та медицині, використання моделей на основі клітинних автоматів у нанотехнологіях. Крім того, клітинні автомати за визначенням є паралельними структурами і тому використовуються для вирішення проблеми моделювання дискретних паралельних процесів. Прикладом застосування принципу клітинних автоматів є гра Джона Конвея «Життя» [2].Ще відома як «гра без гравців» [3].

Особини в цій популяції представлені клітинами в стані 1, тоді як клітини в стані 0 представляють порожній простір.Мірою часу є зміна поколінь колонії, яка відбувається за відомими правилами. Популяція "живих" клітин може рости безперервно, постійно змінюючи своє положення, форму і кількість клітин. Однак у багатьох випадках колонія з часом стає статичною або циклічно повторює один і той самий кінцевий стан. Все це можна використовувати для моделювання відновлення ресурсів у розподілених інформаційних системах.

Мета цієї роботи – розробка і реалізація клітинно-автоматної комп'ютерної моделі для дослідження поведінки структури в умовах знакозмінної зміни параметрів, що визначають ресурсність.

Основна частина

Клітка живого організму має ресурс по числу поділок. Даний ресурс витрачається в процесі життєдіяльності організму. Цікавим є для вивчення засобами комп'ютерного моделювання обмежень життєдіяльності багатоклітинного організму, пов'язаних з витрачанням даного ресурсу. Очікуване застосування результатів моделювання пов'язане з підтриманням оптимальних режимів функціонування систем, до складу яких входять об'єкти живої природи. Важливим прикладним аспектом може виявитися розробка на основі отриманих модельних результатів – від конкретних методик і практичних рекомендацій стосовно екології до оптимізації людської господарської діяльності, а так само до медицини.

Розглянемо ескізно базові принципи організації клітинної структури об'єктів живої природи. В основі життя, в основі всіх живих організмів лежать клітини. Даний факт є зараз емпірично встановленим в біології, в силу чого є вихідним пунктом у побудові біології як науки. Клітини є «цеглинками», з яких складені більш складні структури, що в сукупності і утворює життя в біологічному сенсі. Важливою характеристикою клітини є наявність в ній спадкової інформації у вигляді нуклеїнової кислоти - ДНК. Також визначення відображає найважливішу рису будови клітини: наявність зовнішньої мембрани (плазмолеми), що розмежовує клітку і навколишнє її середовище. Розглянутий процес цікавий як об'єкт моделювання. Цікавим є вивчення засобами комп'ютерного моделювання обмежень життєдіяльності багатоклітинного організму, пов'язаних з витрачанням його ресурсу.

Комп'ютерне (дискретно-математичне) моделювання дозволяє істотно економити ресурси (матеріальні та інтелектуальні), що витрачаються в процесі пізнання і освоєння природи, зокрема в ході розробки прикладних технічних систем. Клітинні автомати дозволяють абстрагуватися від багатьох несуттєвих рис даного об'єкту, зосереджуючи переважну увагу на циклічності процесів, просторовій розподіленості об'єктів, дискретності взаємної незалежності подій. В ресурсних клітинних автоматах додатково береться до уваги тимчасова обмеженість процесів: скорочення можливостей для подальшого їх продовження. Сукупність зазначених якостей створює формальне причинно-наслідкове середовище, зручне для розгляду досліджуваних об'єктів. Основний напрямок дослідження клітинних автоматів - алгоритмічна розв'язність тих чи інших завдань. Також розглядаються питання побудови початкових станів, при яких клітинний автомат буде вирішувати задану задачу.

Цікавий спеціальний клас клітинних автоматів, у яких на кожному кроці еволюції клітинного автомата значення осередку дорівнює якомусь цілому числу (зазвичай обирається з кінцевої безлічі), а новий стан клітини визначається сумою значень клітин-сусідів і, можливо, попереднім станом клітини. Стан клітини на новому етапі залежить від її попереднього стану. В цілому - клітинний автомат є інструмент моделювання. Тому йому, як і кожному інструменту, слід відповідати предмету, об'єкту і ситуації застосування. Тому правомірні, допустимі і доцільні модифікації клітинних автоматів. Головне - щоб вони найкращим чином відповідали області застосування - об'єкту моделювання.

Стосовно до моделювання біологічних об'єктів нами використовуються ресурсно-обумовлені (далі просто ресурсні) клітинні автомати, що дозволяють відслідковувати зміну (витрачання) певних значень параметрів окремих елементів.

Ресурсна клітинно-автоматна модель

Перед початком розробки моделі обговоримо вимоги до неї. Ці вимоги повинні бути пов'язані з практикою експлуатації моделі. Тобто вони обумовлені характером, обсягом і заходом ілюстративності інформації, яка передбачається до отримання засобами моделі, що розробляється. Графічний інтерфейс не потрібен. Надмірність графічного інтерфейсу визначається тим, що інформація виводиться виключно в цифровій формі, в текстовому форматі, безпосередньо в Протокол експерименту. Статистична обробка результатів декількох запусків моделі також не потрібна. У даній роботі ми обмежуємося лише ілюстративної стороною. Достатньою є побудова гістограм розподілу числа клітин КА за значенням їх ресурсу, по декілька запусків моделі. Сама побудова графіків гістограм в рамках моделюючої програми так само не потрібна. Моделююча програма повинна видати числові дані, а графіки будуватимуться в табличному процесорі (Excel, Libre Office або Open Office) при розборі і аналізі результатів, тобто не в єдиному потоці з роботою моделі. Даний підхід з рознесенням етапів моделювання та обробки і візуалізації результатів обрано як максимально зручний з точки зору трудовитрат на розробку і налагодження програми, а так само з урахуванням зручності процесу проведення досліджень. Подібний підхід так само зручний тим, що не пред'являється ніяких істотних вимог до мови програмування та обмежень на стиль програмування. В ескізному (первинному) варіанті програма може бути написана в простому процедурному стилі. Якщо підхід виявиться в цілому вдалим, а програма виявиться як необхідна - для швидкості її роботи і зниження ресурсоємності доцільний функціональний стиль програмування. У цій реалізації вимога функціонального стилю не висувається. Оскільки, як зазначено, можлива подальша переробка програми з виконанням її в функціональному стилі - доцільно спочатку реалізовувати її на мові Python. Ця вимога обумовлена тим, що мова Python підтримує функціональний стиль (в ньому є набір відповідних операторів), і, таким чином, програма з'явиться напрацюванням для подальшого модифікування.

Також необхідно зазначити модельні обмеження. Вони в основному обумовлені мірою абстрактності, яка спочатку закладається в модель. Позитивна сторона наявних обмежень - відсікання факторів і характеристик, що не торкаються на даному (поточному) етапі моделювання. Розглядається (моделюється) тільки двовимірний клітинний автомат. Якщо буде потрібно багатовимірне розширення, товоно може бути завжди доопрацьовано при наявності двовимірного аналога. Але трудомісткість розробки, а також самого процесу роботи з багатовимірної моделлю - зростає в квадратичній або навіть кубічній залежності. Тому, доцільністю диктується двовимірне обмеження. Розглядаються тільки квадратні поля КА. У відношенні реалізації дане обмеження саме несуттєве. Але характерно, що їм задається конкретність. Якщо прийняти поле прямокутної форми, то принаймні це ще один додатковий параметр, за яким потрібно проводити дослідження. Розглядаються варіанти сусідства 4, 6 і 8. Вони є досить простими. В принципі, можуть бути вказані інші значення функції сусідства, але це знову призводить до збільшення параметрів моделі.

Модель передбачається для дослідження на комп'ютерах загального призначення. Спеціальні багато-

процесорні розпаралеленовані системи використовувати не передбачається. Відповідно, в текст програми відповідні (зокрема багатопотокові) можливості не закладаються. Наступне питання пов'язане з використанням будь-яких спеціалізованих математичних пакетів. Модель виконується в самодостатньому варіанті, на стандартній мові Python без будь-яких спеціальних мовних розширень. Уявімо алгоритм моделі у вигляді покрокового опису (рис. 1).

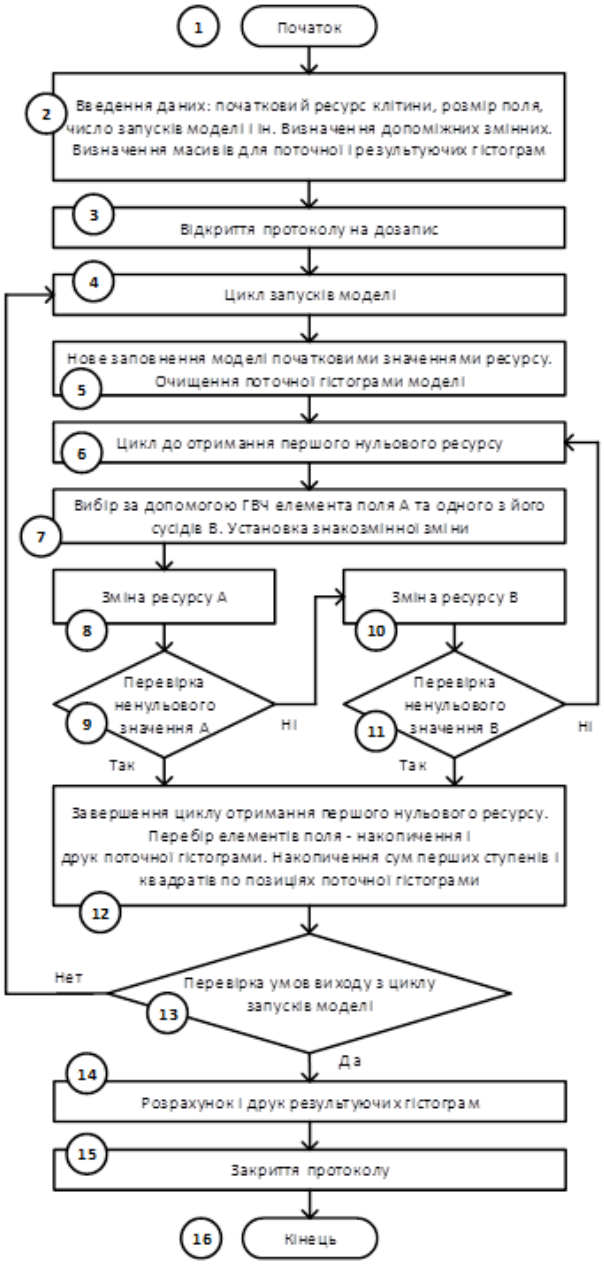


Рис. 1. Структурна схема ресурсної клітинно-автоматної моделі

Крок 1. Пуск. Запуск програми. Параметри запуску (вихідні дані) зберігаються в окремому файлі - стартовому протоколі, який повинен бути доступний при запуску програми. Це робиться для того, щоб програму, що реалізовує модель, можна було запускати автономно, не коригуючи при цьому текст програми (з подальшою перекомпіляцією в виконуваний код) всякий раз при введенні нових початкових даних.

Уточнення початкового тексту програми може знадобитися тільки для модифікування (наприклад, розширення набору параметрів) моделі. Подібне модифікування - є переналагодження моделі для нового дослідження, по суті створення нової моделі. Принаймні, мова може йти про новий (інший) план експерименту, що включає, можливо, нові параметри і (або) нову послідовність операцій.

Крок 2. Введення N кількості застосувань моделі та інших значень параметрів. Ініціалізація експериментального протоколу.

Число N реалізацій моделі є число послідовних перезапусків моделі (всередині одного запуску програми) для набору статистично достовірної вибірки результатів. Важливими параметрами, що вводяться під час запуску програми, є так само конфігурація (прямокутна) і розмір поля клітинного автомата. Конфігурація задається вказівкою розмірів (довжини і ширини) поля. Вся ця інформація вводиться в програму з вхідного (стартового) Протоколу. Протокол експерименту (вихідний Протокол) створюється на дозапис, тобто з параметром append. Ініціалізація Протоколу полягає в наступному: ім'я вихідного Протоколу зчитується з вхідного (стартового) Протоколу. Якщо програма не знаходить файлу з таким ім'ям - вона створює порожній файл. Потім програма відкриває цей файл (колишній перш, знайдений в директорії, або новостворений) на дозапис, без видалення колишнього контенту. Файл залишається відкритим весь час роботи програми і закривається тільки при завершенні (Крок 14 - Стоп). Таким чином, в єдиний Протокол експериментів збирається вся вихідна інформація: не тільки послідовні запуски моделі всередині одного старту програми, а й всі послідовні запуски програми. Крім того, при ініціації Протоколу, в Протокол заносяться всі дані зі стартового протоколу. Таким чином, кожного разу при запуску програми протоколюється повний початковий стан серії з N реалізацій моделі.

Крок 3. Запуск чергової реалізації моделі. Установка початкових значень ресурсів клітин КА. Обнулення гістограми

Як зазначалося, модель всередині програми запускається N-кратно. Тобто після першого проходження моделі, сюди в цю точку Крок 3 буде ще (N-1) повернень. Всякий раз заново встановлюються значення ресурсів клітин клітинного автомата. Ресурс клітини є число її подальших поділів в процесі функціонування організму. У початковому стані (при «народженні» модельного організму) всі клітини КА мають однаковий максимальний ресурс, значення якого задається в стартовому протоколі. При першому проході моделі - масив гістограми ще порожній. Після завершення кожної реалізації в масиві міститься розподіл числа клітин по їх залишковим ресурсам. Як зазначалося, отримання цього розподілу - мета моделювання. Після скидання гістограми в Протокол експерименту, ця інформація всередині програми більше не потрібна. Тому при кожному новому запуску реалізації моделі, масив гістограми обнуляється.

Крок 4. Початком наступного кроку в клітинному автоматі є збільшення лічильника робочого циклу з КА на одиницю. При цьому здійснюється

(моделюється, відтворюється) черговий поділ клітин. Для цього обирається деяка клітка А з поля клітинного автомата (Крок 5) і один з її сусідів В (Крок 6).

Крок 5. Вибір клітини А здійснюється псевдовипадковим чином, тобто позиція клітини в полі КА вказується з використанням генератора випадкових чисел (ГВЧ). Моделювання (в даній версії моделі) здійснюється до першого обнулення ресурсу клітини. Тому процедура вибору клітини А - «спрацьовує» з першого разу і не припускає повторного звернення до ГСЧ. Таким чином, модель «не гальмуватиме» у міру вироблення ресурсу КА.

Крок 6. Вибір В як сусіда А відбувається так само за допомогою ГВЧ. Кількість сусідів є параметром моделі, який вводиться з протоколу ініціалізації.У даній роботі нами розглядаються прості випадки – двовимірні КА з функцією сусідства 4, 6 і 8. Значення 6 відповідає гексагональній (шестикутній) симетрії, значення 4 і 8 - тетрагональній (квадратній), без урахування і з урахуванням діагональних (кутових) квадратів відповідно.

Крок 7. Зменшення на одиницю ресурсів А і В. В даній моделі мається на увазі наступна ситуація: клітка А гине, а одна з клітин її оточення (сусідства), а саме клітина В, ділиться і заповнює вакансію. В результаті цього поділу обидві клітини (клітина В і її клон, що заповнює місце клітини А) втрачають по одній одиниці значення свого ресурсу. Тобто, вони стають «старіше» на одну одиницю (на один «тік внутрішнього годинника») «індивідуального біологічного часу життя», хід якого - в зворотному відліку - ми по суті і відтворюємо в цій моделі.

Крок 8. Моніторинг поля КА. Заповнюється масив гістограми. Проходиться область СА і виділяються клітинки СА. Підраховується кількість комірок у кожному "віці". Знову заповнюється масив гістограми. По суті, масив гістограми щоразу обнуляється і перезаповнюється на даному Кроці алгоритму моделі.

Крок 9. Вивід гістограми в Протокол. Після завершення польового обстеження КА масив гістограми містить поточне розташування комірок КА за віковими групами. Масив скидається до протоколу. Цінність цієї інформації - поточний стан моделі. Інформація може бути корисна якщо ми хочемо простежити «динаміку старіння» КА.

Крок 12. Перевірка досягнення порогового значення витрачання ресурсів КА. В даному варіанті моделі граничним значенням (сигналом для завершення чергової реалізації моделі) є поява першої клітини з нульовим ресурсом. У конкретних додатках, при моделюванні поведінки конкретних біологічних організмів, можливо, критерій завершення буде інший. У інтерпретаційному плані, перша клітина з нульовим ресурсом може бути тільки сигналом до початку редукції певного організму. Якщо моделюється інша ситуація, - критерій завершення роботи моделі може бути інший.

Крок 13. Проводиться запис до протоколу про завершення поточної реалізації моделі. Додатково, в програму може бути вставлена окрема статистична обробка. В даному варіанті такої обробки немає. Завершенням роботи моделі є фіксація фінального стану гістограми.

Крок 14. Перевірка відпрацювання необхідного числа N реалізацій моделі. Як зазначалося вище, під час кожного запуску програми послідовно обробляється N повних шляхів моделі. На даному етапі перевіряється факт відпрацювання всіх (необхідного числа) реалізацій.

Крок 15. Закриття Протоколу. На даному Кроці може реєструватися тривалість роботи програми. Фіксація тривалості може бути цікавою, якщо вивчаються можливості практичної роботи з моделлю, зокрема залежність від розмірів поля КА і з урахуванням конкретної продуктивності обчислювальних коштів.

Крок 16. Стоп. Зупинка програми. Результат роботи програми - Протокол - зберігається в окремому файлі.

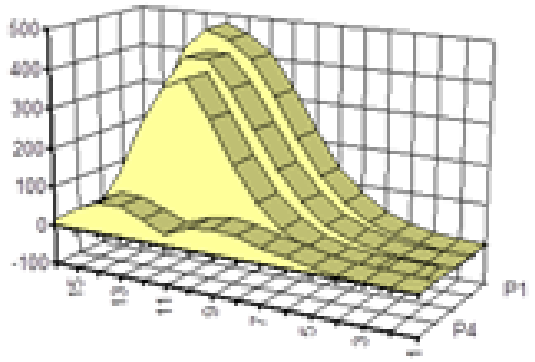
Моделювання та аналіз результатів

З використанням розробленого програмного продукту згідно розробленого плану машинних експериментів, була проведена серія запусків моделі розподіленої системи - нерегулярного клітинного автомата. Мета машинних експериментів (крім демонстрації працездатності розробленого програмного продукту) - отримання оціночних значень тривалості функціонування і динаміки старіння «віртуального організму», відтвореного моделлю.

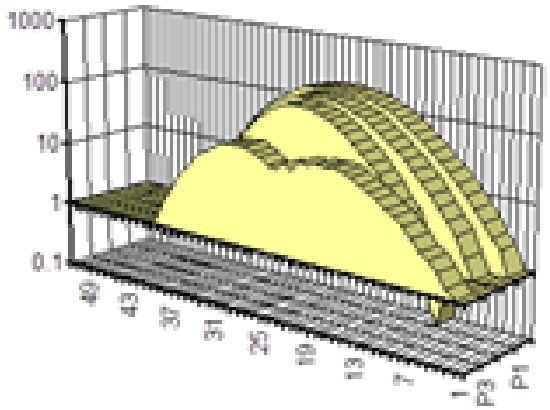
На початку протоколу вказаний тип (топологія) модельованого клітинного автомата (одношаровий двовимірний ресурсний), граничне число реалізацій в кожному запуску моделі (1000), а так само два параметри моделі (ресурс клітини та розмір поля), повний перебір комбінацій яких складає предмет машинного експерименту. Було задано 9 значень ресурсу клітини (10, 15, 20, 25, 30, 35, 40, 45 і 50), а також 4 значення для розміру клітинного поля (50, 100, 150 і 200). Таким чином, в ході машинного експерименту розглядаються поля розміром (50 х 50), (100 х 100), (150 х 150) і (200 х 200). В результаті, сумарне число запусків моделі - 36.

Граничне число реалізацій - 1000 - вибрано як компромісне. Критерій компромісу: при такому числі забезпечується отримання змістовних результатів за прийнятний час на наявних типових обчислювальних засобах (персональних комп'ютерах). Дані після нормування округлені до цілих значень. У такому вигляді вони більш наочно розміщуються в таблиці. Для полегшення візуального аналізу за таблицями побудовані графіки. Кожен графік показує 15 розподілів числа клітин поля за значенням ресурсу. Деякі з них представлені в роботі (рис. 2).

Слід звернути увагу на обмеження контексту експерименту і вплив їх на отримані результати. Так, зазначена вище обмеженість тимчасових рамок і її вплив на стійкість результатів - чітко видно на зіставленні графіків в лінійному і логарифмічному масштабах. Особливо цікавий «хвіст» графіка, відповідний «періоду старості віртуального організму», наближенню ситуації появи першого нульового ресурсу. Мабуть, цікавим напрямком вивчення може з'явитися підбір параметрів моделі, при якому росте гладкість зазначеного «хвоста».



а – ресурс клітини 15



б – ресурс клітини 45

Рис. 3. Результати моделювання (розмір поля 50)

«Гладкий» та «тривалий» хвіст графіка - вказівка на більш високу рівномірність використання ресурсу осередків клітинного автомата.

Інтерпретація параметрів предметної області може забезпечити при цьому цілком конкретні позитивні рекомендації.

Висновки

Розроблені і реалізовані клітинно-автоматні комп'ютерні моделі для дослідження поведінки структури в умовах знакозмінної зміни параметрів, що визначають ресурсність.

Розроблено програмне забезпечення для реалізації моделі; продемонстрована працездатність програми; побудовані і виконані пробні плани машинних експериментів.

Дані короткі обґрунтування до вибору базових програмних засобів для реалізації моделі.

Відзначено, що основне призначення моделі - отримання даних.

У зв'язку з цим визнано доцільне рознесення інформаційних і ілюстративно-оформлювальних аспектів моделювання за рахунок використання різних спеціалізованих програмних продуктів.

Конкретизовано поділ функцій моделі по програмним продуктам.

Описано основні етапи налагодження моделі з використанням протоколів.

Також в статті прокоментована графічна візуалізація отриманих результатів.

СПИСОК ЛІТЕРАТУРИ

1. Михаль, О.Ф. Глобально-исторический контекст развития средств вычислительной техники // Бионика интеллекта. - 2014. – №1 (82). – С. 55–62
2. Gardner, Martin (October 1970). "The fantastic combinations of John Conway's new solitaire game 'life'" (PDF). Mathematical Games. Scientific American. Vol. 223, no. 4. pp. 120–123.
3. Berlekamp, E. R. Winning Ways for your Mathematical Plays / E. R. Berlekamp, John Horton Conway, R. K. Guy. – 2nd. – A K Peters Ltd, 2001–2004.

Received (Надійшла) 02.12.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Modeling of resource coordination of distributed information systems

D. Diachenko, A. Huk, O. Mikhal

Abstract. Topicality. The paper analyzes the problems associated with modeling the behavior and interaction of objects of living nature in biological systems. The relevance of developments in this direction of modeling, which is at the interface between biology and the general theory of systems, is noted. The range of issues to be modeled within the framework of this work has been specified. The cell of a living organism has a resource in the number of divisions. This resource is consumed in the process of vital activity of the organism. It is interesting to study by means of computer modeling the limitations of the vital activity of a multicellular organism related to the consumption of this resource. Computer modeling allows you to significantly save resources that are spent in the process of learning and mastering nature, in particular, in the course of developing applied technical and information systems. Cellular automata make it possible to abstract from many non-essential features of objects, focusing mainly on the cyclicity of processes, the spatial distribution of objects, the discreteness of mutual independence of events. In resource-based cellular automata, the temporal limitation of processes is additionally taken into account: the reduction of possibilities for their further continuation. The set of these qualities creates a formal cause-and-effect environment, convenient for examining the objects under study. A resource is the value of a certain parameter (or several parameters), critical for the continued functioning of a certain element of a cellular automaton. The resource can be set initially and can decrease during the functioning of the cellular automaton. The requirements for the developed model are formulated, which are determined by the nature, volume and degree of illustrativeness of the information that is expected to be obtained. The execution language of the model is Python. The algorithm of the model is presented in the form of a step-by-step description, with the necessary commenting, which is used later in the development of the software implementation of the model. The structural diagram of the model is presented and commented. The correspondence of the blocks of the structural diagram of the step-by-step description of the algorithm of the model is shown. The purpose of the work is the development and implementation of a cellular-automatic computer model for studying the behavior of the structure under conditions of sign-variable change of the parameters that determine the resource capacity.

Keywords: cellular automaton, resource, distributed information system, resource recovery, protocol.

УДК 351.865 **doi: 10.26906/SUNZ.2024.1.066**

Є. О. Живило, І. В. Ромашко

Національний університет «Полтавська політехніка імені Юрія Кондратюка», Полтава, Україна

ПРОТОКОЛ СПІЛЬНИХ ДІЙ СУБ'ЄКТІВ ЗАБЕЗПЕЧЕННЯ КІБЕРБЕЗПЕКИ ПІД ЧАС РЕАГУВАННЯ НА КІБЕРІНЦИДЕНТИ, А ТАКОЖ ПРИ УСУНЕННІ ЇХ НАСЛІДКІВ

Анотація. Кіберпростір разом з іншими фізичними просторами визнано одним з театрів воєнних дій. Набирає сили тенденція зі створення кібервійськ (Дорожня карта створення Кібервійськ Збройних Сил України – наказ Генерального штабу Збройних Сил України, від 22.04.2022 №48) до завдань яких належить не лише забезпечення захисту критичної інформаційної інфраструктури держави від кібератак, а й проведення превентивних наступальних кібердій (проведення кібероперацій) у кіберпросторі, що включає порушення сталого функціонування критично важливих об'єктів інфраструктури противника шляхом руйнування електронно-комунікаційних систем, які управляють такими об'єктами. Прогнозується зростання інтенсивності міждержавного протиборства і розвідувально-підривної діяльності у кіберпросторі. Розширюється коло держав, які намагаються сформувати власну кіберрозвідку, оволодіти сучасними технологіями розвідувально-підривної діяльності у кіберпросторі, посилюють державний контроль за національними сегментами мережі Інтернет. Зважаючи на досвід ведення бойових дій під час введення правового режиму воєнного стану та враховуючи невизначеність суб'єктів та об'єктів, їх функцій та завдань для дій в певних сферах, в тому числі і у сфері кібербезпеки, в мирний час, призвів до незлагодженості та неузгодженості цих дій суб'єктами забезпечення кібербезпеки держави. А враховуючи, що з введенням правового режиму воєнного стану певні суб'єкти міняють своє місцезнаходження, переміщують інформаційні активи та обладнання на нові місця дислокації з використанням хмарних сервісів, зазначене доволі сильно ускладнює процес узгодження та координації дій щодо реагування на кіберінциденти, а також усунення їх наслідків. Це призводить до вимушеного перерозподілу завдань та функцій по виконанню заходів кіберзахисту на різних об'єктах. За цих умов, на постійній чи тимчасовій основі створюються нові суб'єкти кіберзахисту, що потребує часу на набуття ними спроможностей для виконання завдань за призначенням. У такій ситуації Україна має бути здатною забезпечити свій соціально-економічний розвиток у цифровому світі, що вимагає набуття спроможності ефективно стримувати деструктивні дії в кіберпросторі, сталого реагування на загрози в кібернетичному просторі, досягнення кіберстійкості на всіх рівнях та взаємодії складових сектору безпеки і оборони щодо забезпечення кібербезпеки в рамках кібероборони держави. Отже, виходячи з необхідності наукового обґрунтування інституційних засад постає необхідним чітко визначити: "перелік суб'єктів забезпечення кібербезпеки щодо виконання дій, встановлених цим Протоколом", як в мирний час так і в умовах правового режиму воєнного стану; зазначеним вище суб'єктам їх роль та місце, перелік та порядок дій під час реагування на кіберінциденти та усунення їхніх наслідків, як в мирний час так і в умовах правового режиму воєнного стану. При цьому, наукова новизна очікуваних результатів полягає в теоретичному обґрунтуванні та наданні практичних рекомендацій щодо вдосконалення механізмів управління та взаємодії складовими (х) сектору безпеки і оборони під час планування підготовки держави до кібероборони, проведення заходів з нейтралізації та активної протидії кіберзагрозам в національному сегменті кіберпростору держави.

Ключові слова: суб'єкти забезпечення кібербезпеки, кіберпростір, кіберзахист, активні кібердії, деструктивні кібератаки, критична інформаційна інфраструктура.

Постановка проблеми у загальному вигляді

Сьогодні, створення умов безпечних спільних дій суб'єктів забезпечення кібербезпеки (далі – СЗК) в національному сегменті кіберпростору України, реалізація державно-приватної взаємодії у сфері кібербезпеки (далі – КБ), а також їх об'єднане застосування в інтересах особи, суспільства і держави є доволі суттєвим та змістовним завданням.

За цих умов посилення спроможностей суб'єктів сектору безпеки та оборони для забезпечення ефективної боротьби із кіберзагрозами воєнного характеру, кібершпигунством, кібертероризмом та кіберзлочинністю, поглиблення міжнародного співробітництва у цій сфері, забезпечення кіберзахисту державних електронних інформаційних ресурсів, інформації, а також інформаційної інфраструктури, яка знаходиться під юрисдикцією України набуває першочергового значення і стає запорукою подальшого успіху на шляху створення безпеки життєво важливих національних інтересів України у кіберпросторі.

Відповідно до статті 5 закону України "Про основні засади забезпечення кібербезпеки України" координацію діяльності у сфері КБ як складової національної безпеки України здійснює Президент України через очолювану ним Раду національної безпеки і оборони України (Далі – РНБО України). В свою чергу Національний координаційний центр кібербезпеки (Далі – НКЦК), як робочий орган РНБО України здійснює координацію та контроль за діяльністю суб'єктів сектору безпеки і оборони які забезпечують КБ.

Внаслідок повномасштабної збройної агресії росії проти України функціонування національної системи кібербезпеки (Далі – НсКБ) під час дії правового режиму воєнного стану в Україні було частково нівельовано.

Тому з метою визначення правових та організаційних основ забезпечення національних інтересів України у кіберпросторі було запропоновано внести

зміни до законів України:

- "Про Національний банк України" де визначався порядок функціонування, вимоги та заходи із забезпечення кіберзахисту та інформаційної безпеки у банківській системі України;

- "Про оборону України" щодо здійснення заходів з кібероборони (активного кіберзахисту) (Далі – КО) для захисту суверенітету держави та забезпечення її обороноздатності, запобігання збройному конфлікту та відсічі збройній агресії.

- "Про Державну службу спеціального зв'язку та захисту інформації України" щодо захисту у кіберпросторі державних інформаційних ресурсів та інформації, координації діяльності СЗК щодо кіберзахисту [8], впровадження організаційно-технічної моделі кіберзахисту, координація, організація та проведення аудиту захищеності комунікаційних і технологічних систем об'єктів критичної інфраструктури (Далі – ОКІ) на вразливість.

Паралельно з цим, також слід відмітити січневі події 2022 року, які відбулись на тлі загострення ситуації в Україні. Так, у ніч з 13 на 14 січня низка українських урядових ресурсів зазнала атак хакерів, які здійснили Deface (англ. deface – заміна сторінок сайтів) і, як повідомили експерти Microsoft, запустили шкідливе програмне забезпечення, замасковане під програми-вимагачі.

При цьому необхідно зауважити, що співробітники НКЦК діяли відповідно до своїх повноважень, визначених Положенням, затвердженим Указом Президента України від 7 червня 2016 року № 242, а саме здійснювали координацію та контроль за діяльністю суб'єктів сектору безпеки та оборони, які забезпечували кібербезпеку, а також аналізували дані про кіберінциденти щодо державних електронних інформаційних ресурсів та критично важливих об'єктів інфраструктури держави [1].

Зважаючи на ситуацію яка склалась, Апаратом РНБО України було прийнято рішення, щодо виконання ряду заходів які дозволять запобігти (унеможливлять) в майбутньому ймовірним кіберінцидентам [5], а саме:

1. Забезпечення розробки та впровадження узгодженого Протоколу спільних дій СЗК, власників (розпорядників) об'єктів критичної інформаційної інфраструктури (Далі – ОКІІ) при виявленні, попередженні, припиненні кібератак та кіберінцидентів, а також під час усунення їхніх наслідків;

2. Запровадження механізмів додаткового стимулювання мотивації до праці фахівців сектору безпеки та оборони, які беруть безпосередню участь в організації та реалізації заходів щодо протидії кіберзагрозам;

3. Активізація співпраці із закордонними партнерами щодо протидії кібератакам на критичну інформаційну інфраструктуру, проведення розслідувань таких кібератак, встановлення причин та умов, що призвели до їх скоєння.

Отже враховуючи зазначене вище, завдання щодо розробки Протоколу спільних дій СЗК під час реагування на кіберінциденти, а також при усуненні їх наслідків постає сама собою.

Аналіз останніх досліджень і публікацій

Активність у кіберпросторі України має систематичний характер. Протягом першого півріччя 2022 року Україна знову опинилася в центрі кібератак, спрямованих на її критичну інфраструктуру (Далі – КІ). Сьогодні протистояння в кіберпросторі національного сегменту відбувається на тлі військової агресії з боку росії.

Для реалізації її планів, до оперативного складу угрупування діяльність якого спрямована на реалізацію активних дій у національному кіберпросторі України залучені такі російськомовні хакерські угруповання, як ФАПСІ, КиберБеркут, Iridium, Sofacy Group (ГРУ ГШ РФ), APT29 (ФСБ РФ), Turla (ФСБ РФ), UAC-0010 (ФСБ РФ), Sandworm, Тролі з Ольгіна та інші.

При цьому, на противагу кібер угрупованням росії розгорнута ціла українська ІТ-армія яка нараховує понад 300 тисяч кіберфахівців. До її складу входять: Служба безпеки України (Далі – СБ України), CERT-UA Державної служби спеціального зв'язку та захисту інформації України (Далі – ДССЗЗІ України), InformNapalm, IT Army of Ukraine, Український Кібер Альянс (FalconsFlame, Trinity, Ruh8), Українські Кібер Війська, Центр "Миротворець" та інші. Вона об'єднує українських та міжнародних ІТ-фахівців, засновників, творців, комунікаторів для боротьби з російською агресією на кіберфронті. В цих умовах ІТ-армія проводить кібератаки і DDoS-атаки на ресурси бізнес-корпорацій ("Газпром", "Лукойл"), банків ("Сбербанк", ВТБ, "Газпромбанк"), а також на сайти держ-служб росії, кремля й держдуми.

Попри зазначені перемоги все ж таки їхні атаки на наші критичні системи, об'єкти та інфраструктуру все ж таки мали (ють) успіх.

За оприлюдненими даними Урядової команди реагування на комп'ютерні надзвичайні події України CERT-UA до групи загроз [10], яка створила реальну небезпеку вчинення актів кібертероризму та кібердиверсій стосовно національної інформаційної інфраструктури за цей період слід віднести: WhisperGate /WhisperKill, FoxBlade, він же Hermetic Wiper, SonicVote, HermeticRansom, CaddyWiper, DesertBlade, Industroyer2, Lasainraw, IssacWiper, FiberLake, DoubleZero, Cobalt Strike Beacon та інші [12].

Виходячи з використаних тактик противника, слід зазначити, що в основному вони були спрямовані на:

блокування роботи ОКІІ, телекомунікаційного обладнання та приладдя органів державної влади, доведення його до непридатності для використання з наступним виведенням його з ладу взагалі;

застосування фішингових атак на держслужбовців, військових, працівників КІ з метою отримання конфіденційної або секретної інформації, а також фінансових даних;

отримання кіберзловмисниками інформації щодо облікових даних, які дозволяють дістати доступ до систем, сервісів або служб, якими користу-

ються відповідні співробітники, персонал, особи та працівники на своїх персональних ЕВМ на об'єктах інформаційної діяльності установ (організацій) державної, приватної та колективної форми власності;

розповсюдження в популярних мессенджерах шахрайських публікацій щодо отримання соціальних виплат, фінансової допомоги та інше, метою яких є компрометація та отримання персональної інформації власників платіжних карток.

Але попри зазначені негаразди в цих умовах відбулась згуртованість дій фахівців КБ Європейського Союзу, європейських країн, уряду США, НАТО і ООН дії яких спрямовані на противагу деструктивним атакам, шпигунським операціям, руйнування чи деградацію української мови, обмеження урядових та військових функцій і підривання довіри громадськості до цих же інституцій.

В цих умовах урядом України проводиться ряд заходів пов'язаних з розбудовою державної системи захисту КІ, визначаються правові та організаційні засади забезпечення її діяльності, реалізується державна політика у сфері захисту КІ, безпечним користуванням українцями телефонами, інтернетом та інше.

Аналізуючи досвід реагування на кіберзагрози та кіберінциденти спеціалістами провідних країн світу в галузі КБ, українські фахівці з КБ дійшли висновку, що велика кількість таксономій і схем класифікації інцидентів забезпечують чудові вказівки в рамках роботи центру безпеки (SOC) одного підприємства, установи або організації.

Однак такі системи не розглядають визначення пріоритетів інцидентів або оцінку ризиків із загальнонаціональної точки зору, що може залучати велику кількість різноманітних підприємств. Великі національні операційні центри з КБ, як-от Агентство кібербезпеки та безпеки інфраструктури (CISA), повинні оцінювати ризики, вміщуючи різноманітну групу власників та операторів приватних критично важливих інфраструктурних об'єктів, а також відомств і агенцій уряду США.

Національна система оцінки кіберінцидентів (NCISS) розроблена, щоб забезпечити повторюваний і послідовний механізм для оцінки ризику інциденту в цьому контексті.

При цьому необхідно наголосити, що NCISS базується на Спеціальній публікації Національного інституту стандартів і технологій (NIST) 800-61 Rev. 2, Посібнику з обробки інцидентів з комп'ютерною безпекою, і розроблено для включення категорій потенційного впливу на конкретні суб'єкти, які дозволяють персоналу CISA оцінювати серйозність ризику та пріоритет інцидентів з загальнонаціональної точки зору. NCISS дозволяє подібному інциденту, який зазнали дві різні зацікавлені сторони, мати суттєво різну оцінку на основі потенційного впливу кожного постраждалого суб'єкта на національному рівні. Система не призначена для абсолютного оцінювання ризику, пов'язаного з інцидентом.

Так NCISS використовує середнє арифметичне зважене, щоб отримати оцінку від нуля до 100. Ця оцінка керує процесами сортування та ескалації інцидентів CISA і допомагає визначити пріоритетність обмежених ресурсів реагування на інциденти та необхідний рівень підтримки для кожного інциденту.

Наразі система не розроблена для підтримки випадків, коли кілька взаємопов'язаних інцидентів можуть збільшити загальний ризик, наприклад, кілька одночасних компромісів організацій у певному секторі чи регіоні.

Однак подібні події все ще можуть бути легко посилені за допомогою експертного втручання людини.

Вхідні дані для системи оцінювання є сумішшю дискретних та аналітичних оцінок. Хоча всі спроби звести до мінімуму індивідуальні упередження за допомогою тренувань і вправ, різні індивідуальні рахунки неминуче матимуть дещо різні погляди на свої відповіді на деякі запитання щодо оцінки.

Використання кількох дискретних вхідних даних, які можна перевірити, зменшує вплив будь-якого окремого аналітичного фактора, підвищуючи загальну надійність системи.

NCISS узгоджується зі схемою серйозності кіберінцидентів (CISS), щоб рівні серйозності в NCISS відображалися безпосередньо на рівнях CISS.

Виділення невирішених раніше частин загальної проблеми

За результатами опрацювання існуючої нормативно-правової бази держави, проєктів нормативних документів та технічних рішень щодо роботи центрів безпеки СЗК, які визначають пріоритети інцидентів або здійснюють оцінку ризиків із загальнонаціональної точки зору слід відмітити наступне.

Вперше Протокол спільних дій реагування на загрози кібербезпеці держави було підготовлено Адміністрацією ДССЗЗІ України на виконання вимог підпункту "г" підпункту 3 пункту 2 Рішення РНБО України від 29 грудня 2016 року "Про загрози кібербезпеці держави та невідкладні заходи з їх нейтралізації", введеного в дію Указом Президента України від 13 лютого 2017 року № 32, та пункту 2 Плану заходів на 2017 рік з реалізації Стратегії кібербезпеки України, затвердженого розпорядженням Кабінету Міністрів України від 10 березня 2017 року № 155-р. Проектну назву Постанови Кабінету Міністрів України було запропоновано наступну – "Про затвердження Протоколу спільних дій основних СЗК, суб'єктів кіберзахисту та власників (розпорядників) ОКІІ під час попередження, виявлення, припинення кібератак та кіберінцидентів, а також при усуненні їхніх наслідків" [2].

Зазначеним Протоколом передбачалось міністерствам, іншим центральним органам виконавчої влади визначити (створити) підрозділи (команди, центри, групи), які забезпечуватимуть кіберзахист та реагування на кіберзагрози щодо ОКІІ у відповідній галузі або сфері діяльності та/або покласти функції з кіберзахисту на підрозділи із захисту інформації (Далі – ЗІ). Також, державним органам, органам місцевого самоврядування, органам управління

Збройними Силами, іншими військовими формуваннями утвореними відповідно до законів, правоохоронним органам, підприємствам, установам та організаціям, у власності чи розпорядженні яких є ОКІІ та/або до сфери управління яких належать (перебувають в управлінні) підприємства, установи та організації, що є власниками (розпорядниками) таких об'єктів [7], пропонувалось організувати створення або створити на таких об'єктах підрозділи кіберзахисту та/або покласти функції з кіберзахисту на підрозділи із ЗІ.

Тож, на той час проєкт Протоколу носив лише загальний характер та встановлював лише перелік взаємно пов'язаних у часі та за цілями обов'язкові дії СЗК під час реагування на кіберінциденти та усунення їхніх наслідків.

При цьому, сам порядок (механізм) здійснення цих спільних дій, роль та місце кожного СЗК визначено не було, що надалі залишало на низькому рівні організацію взаємодії. Крім того, перелік обов'язкових дій СЗК, зазначених в проєкті Протоколу, був загальним та не відображав специфічних завдань та функцій цих СЗК.

Також, в проєкті Протоколу було визначено шість рівнів кіберзагроз, при цьому зазначалось, що заходи реагування на кіберінциденти на всіх етапах виконувались для кожного рівня кіберзагрози. Проте, ймовірніше всього, для кожного рівня кіберзагроз було необхідно здійснювати різні за складністю дії з реагування на кіберзагрози та кіберінциденти, а також ймовірніше всього вони повинні були носити різний ступінь залученості СЗК.

В зазначеному проєкті спільно розглядались функції та завдання Міністерства оборони України та Генерального штабу Збройних Сил України, хоча в положенні і інструкціях зазначених органів управління (органах військового управління) різне призначення, тому з метою уникнення дублювання зазначених функцій під час виконання дій в ході реагування на кіберінциденти та усунення їхніх наслідків, а також враховуючи, що ці функції і завдання є різними, пропонувалось завдання на кожному етапі реагування на кіберінциденти визначати окремо для Міністерства оборони України і окремо для Генерального штабу Збройних Сил України та Збройних Сил України.

Слід відмітити, що зміст проєкту Протоколу не відповідав його назві.

В проєкті Протоколу не було визначено перелік узгоджених за часом та завданнями спільних та взаємопов'язаних обов'язкових дій СЗК під час реагування на кіберінциденти. Були перелічені лише загальні завдання СЗК, визначені іншими нормативно-правовими документами.

Ще доволі суттєвою проблематикою було те, що внаслідок невдалого перекладу англійських назв етапів реагування на кіберінциденти частково втрачалась відповідність назв цих етапів їх змісту.

Тому, назву етапу реагування Containment, що перекладено як "стримування", в подальшому пропонується змінити на "локалізація" (змістом цього етапу є ізоляція уражених елементів об'єктів кіберзахисту з метою нерозповсюдження загрози); назву Eradication, що перекладено як "усунення наслідків", пропонується замінити на "усунення загрози" (так усунення наслідків кіберінцидента здійснюється у ході наступного етапу – "відновлення"). Також, з метою забезпечення взаємосумісності та взаємодії з підрозділами КБ Європейського союзу та НАТО пропонується вказати англійські варіанти назв етапів реагування на кіберінциденти відповідно до NIST Special Publication 800-61.

Для вирішення завдань, щодо врегулювання та імплементації норм та правил міжнародних організацій сфери КБ та КО пропонується дати визначення термінам "критичні інформаційні активи" та "критичні інциденти безпеки", які наведені у пункті проєкту Протоколу, вираз "є інформацією з обмеженим доступом" замінити виразом "розповсюджується відповідно до Загальних правил обміну інформацією про кіберінциденти (Протокол TLP), схвалених рішенням НКЦК при РНБО України (пункт 3.1 Протоколу № 18 від 25.10.2021)", додати: "забезпечення обізнаності користувачів з базових питань кібербезпеки" та "проведення кібернавчань".

Проаналізувавши чинні положення (аксіоматики) існуючої законодавчої, державної та відомчої нормативно-правової бази, а також нормативно-правового поля міжнародних організацій пропонується об'єднати та викласти в редакції нового Проєкту заходи по розробленню детального плану та стандартних операційних процедур реагування на кіберінциденти.

Апробацію зазначеного плану провести під час проведення заходів колективної підготовки СЗК держави та, за необхідності, внести відповідні коригування до них.

Враховуючи неоднозначне тлумачення визначення рівнів кіберзагрози пропонується розглянути можливість визначення рівнів кіберзагрози за моделлю NCISS – National Cyber Incident Scoring System (США), опис якої доступний за посиланням: https://us-cert.cisa.gov/CISA-National-Cyber-Incident-Scoring-System.

Розглядаючи порядок інформування СЗК, який виявив факт проведення протиправних дій в кіберпросторі, інших СЗК, пропонується таке інформування здійснювати через одного із "ключових" суб'єктів НсКБ, наприклад через НКЦК при РНБО України або ДССЗЗІ України. Крім того, вважається недоцільним інформувати всіх без виключення СЗК, вказаних в проєкті Протоколу.

Також пропонується визначити ступені критичності тих кіберінцидентів, про які СЗК інформуються негайно.

В додатках зазначеного проєкту Протоколу пропонується навести завдання НКЦК при РНБО України. Також для кожного з етапів реагування на кіберінциденти з урахуванням рівнів кіберзагроз пропонується розробити алгоритми виконання конкретних практичних спільних дій СЗК під час реагування на кіберінцинденти, а також при усуненні їхніх наслідків. Інформацію, яка наведена у додатку до проєкту Протоколу "Додаткові завдання окремих

суб'єктів взаємодії на кожному етапі реагування на кіберінциденти" пропонується використати для розробки зазначених алгоритмів.

В цілому, вивчаючи проєкт документа слід зауважити, що Протокол встановлює лише "перелік взаємно пов'язаних у часі та за цілями обов'язкових дій СЗК під час реагування на кіберінциденти та усунення їхніх наслідків".

Проте сам порядок (механізм) здійснення цих спільних дій, роль та місце кожного СЗК не визначено, що надалі залишає на низькому рівні організацію взаємодії.

Крім того, перелік обов'язкових дій СЗК, зазначених в проєкті Протоколу, є загальним та не відображає специфічних завдань та функцій цих СЗК.

Виходячи із викладеного, пропонується доопрацювати проєкт Протоколу в частині визначення не тільки переліку, але й порядку (механізму) спільних дій СЗК під час реагування на кіберінциденти та усунення їхніх наслідків, із врахуванням специфічних завдань та функцій цих СЗК.

Формулювання цілей статті (постановка завдання)

В даній статті проаналізовано особливості взаємодії основних суб'єктів НсКБ в умовах:

особливого періоду,

правового режиму воєнного та надзвичайного стану,

здійснення заходів із забезпечення національної безпеки і оборони, відсічі і стримування збройної агресії,

проведення операції об'єднаних сил (антитерористичної операції).

За результатами опрацьованого аналізу, запропоновано перелік взаємно пов'язаних у часі та узгоджених за алгоритмами обов'язкових дій СЗК під час реагування на кіберінциденти та усунення їх наслідків, проведення превентивних наступальних дій (операцій) у кіберпросторі з врахуванням дефініцій та визначень які встановлені спільними міжвідомчими наказами основних суб'єктів НсКБ.

Виклад основного матеріалу дослідження

В рамках виконання вимог плану оборони України, введеного в дію Указом Президента України від 24 лютого 2022 року № 70/2022 "Про рішення Ради національної безпеки і оборони України від 24 лютого 2022 року "Про введення в дію плану оборони України та Зведеного плану територіальної оборони України", та вимог Закону України "Про основні засади забезпечення кібербезпеки України" в частині відбиття воєнної агресії у кіберпросторі (КО) та впровадження заходів із забезпечення кіберзахисту критичної інформаційної інфраструктури СЗК під час реагування на кіберінциденти [3], а також при усуненні їхніх наслідків вважається за потрібне:

1. Уточнити склад СЗК щодо виконання дій, встановлених зазначеним вище документом та визначити основних суб'єктів НсКБ і сил кіберзахисту з метою виконання їх спільних дій під час реагування на кіберінциденти, а також при усуненні їхніх наслідків.

В подальшому пропонується віднести до основних суб'єктів НсКБ наступні елементи:

сили безпеки і оборони, сили кіберзахисту, НКЦК, як робочий орган РНБО України;

центральні органи виконавчої влади, інші державні органи, які забезпечують формування та/або реалізацію державної політики в одній чи кількох сферах, або безпосередньо проводять відповідно до компетенції заходи із забезпечення КБ;

місцеві органи виконавчої влади, органи місцевого самоврядування, що провадять діяльність у сфері ЗІ та кіберзахисту;

ОКІ незалежно від форми власності; підприємства, установи та організації незалежно від форми власності, що провадять діяльність у сфері ЗІ та кіберзахисту, взаємодіють із силами кіберзахисту або виконують роботи та надають послуги за державні кошти [4].

В умовах правового режиму воєнного стану цей перелік може бути змінений Генеральним штабом Збройних Сил України.

2. Використовувати наступні терміни та визначення у новій редакції Протоколу, а саме:

сили кіберзахисту – урядова команда реагування на комп'ютерні надзвичайні події України CERT-UA, інші команди реагування на комп'ютерні надзвичайні події, підрозділи (групи, команда, служби) ЗІ, підприємства, установи та організації незалежно від форми власності, які провадять діяльність та/або надають послуги, пов'язані з кіберзахистом [6];

рівень кіберзагрози – показник небезпеки від потенційного або незворотнього настання кіберінциденту, що може спричинити значні руйнівні для критичної інформаційної інфраструктури країни наслідки;

Інші терміни вживати у значенні, наведеному в Законах України "Про основні засади забезпечення кібербезпеки України", "Про національну безпеку України", "Про захист інформації в інформаційно-телекомунікаційних системах", "Про електронні комунікації".

3. Встановити такі етапи реагування на кіберінциденти: підготовка; виявлення та аналіз; стримування; усунення; відновлення; заходи після інциденту.

При цьому етап підготовки спрямований на забезпечення готовності реагування на кіберінциденти, а також запобігання їм, і передбачає виконання таких заходів:

розробка та оновлення політик безпеки;

затвердження детального плану реагування на кіберінциденти, перевірка використання його та, за необхідністю внесення коригувань до нього, оцінка ризиків;

визначення критичних інформаційних активів, визначення критичних інцидентів безпеки; створення, перевірка, проведення навчань;

створення/або визначення команди реагування на інциденти CSIRT/CIRT/CERT, порядку комунікації з правоохоронними органами, CERT-UA.

Етап виявлення та аналізу передбачає виявлення подій, які можуть спричинити виникненню інциденту, узагальнення інформації щодо них та наявності вразливостей, і передбачає такі заходи:

забезпечується постійний контроль та моніторинг ІТ-систем;

здійснюється виявлення аномалій, виявлення та аналіз, а також підтвердження інцидентів безпеки;

при виявленні інциденту, виконується початковий аналіз його з метою визначення масштабності, причини виникнення і яким чином відбувається інцидент (інструменти або методи, які використовувалися для атаки, якими вразливими місцями скористалися);

відбувається збір додаткових даних з різних джерел, їх дослідження, встановлення типу інциденту, згідно Переліку категорій кіберінцидентів і рівня його критичності;

при аналізі виникнення інциденту команда суб'єкта взаємодії отримує достатньо інформації для визначення наступних заходів, як стримування, усунення та відновлення;

всі зібрані дані документуються, а команда суб'єкта взаємодії інформує про кіберінцидент відповідно до класифікації (таксономії) кіберінцидентів та протоколу обміну інформацією про кіберінциденти.

Етап стримування спрямований на забезпечення розроблення суб'єктами взаємодії цілеспрямованої стратегії її відновлення, а також призначення та реалізацію першочергових заходів стримування для запобігання поширенню загрози.

При здійсненні довгострокового стримування відбувається внесення тимчасових виправлень до систем задля можливості їх застосування до завершення налаштування систем (їх елементів), які відтворені з їх неуражених копій.

Етап усунення наслідку кіберінциденту спрямований на реалізацію заходів з видалення шкідливого програмного забезпечення з усих уражених систем, усунення наслідків впливу інциденту, визначення першопричин іициденту та вжиття заходів для запобігання атакам подібного типу у майбутньому.

Етап відновлення передбачає реалізацію суб'єктом взаємодії заходів з відновлення системи до штатного режиму функціонування та переконання в її стабільному функціонуванні, що передбачає:

підключення раніше ізольованих уражених сегментів після відновлення до основної системи;

вжиття заходів із запобігання додатковим атакам;

тестування, перевірка та контроль відновлених після ураження систем для їх повернення до штатного функціонування з урахуванням встановленого часу для відновлення.

Етап заходів: після інциденту передбачає:

аналіз отриманого досвіду інциденту після його закінчення, проведення навчань та зустрічей з метою обміну досвідом;

перегляд та внесення змін до політик безпеки та документації за результатами дослідження інциденту;

оцінку дій щодо реагування на інцидент з метою покращення процесів реагування на кіберінциденти у майбутньому.

4. В залежності від рівня критичності кіберінцидентів для прийняття рішення щодо впровадження додаткових заходів кіберзахисту встановлюються такі рівні кіберзагроз:

базовий ("білий"),

низький ("зелений"),

середній ("жовтий"),

високий ("помаранчевий"),

серйозний ("червоний"),

надзвичайний ("чорний").

Базовий ("білий") рівень вказує на:

нульовий рівень загроз від настання кіберінцидентів,

наявність несуттєвих подій,

стале функціонування ОКІІ держави.

Низький ("зелений") рівень вказує на низький стан загроз, динаміка критичності якого залежить від настання критичності кіберінцидентів. Не існує жодної незвичайної активності, окрім звичайного занепокоєння про відомі хакерські дії, віруси та іншу зловмисну діяльність.

Середній ("жовтий") рівень вказує на середній рівень загроз від настання кіберінцидентів, при якому спостерігається збільшення хакерських дій, вірусів або іншої зловмисної діяльності.

Існує потенціал для кіберзловмисної діяльності, але виявлена невідома раніше або відома така діяльністъ, але значного впливу на системи не відбулося.

Високий "помаранчевий" рівень вказує на високий рівень загроз від настання кіберінцидентів через зростаючі хакерські дії, віруси або іншу зловмисну діяльністъ, яка компрометує системи або звужує надання послуг.

На цьому рівні існують відомі вразливості, які використовуються з помірним ступенем пошкодження чи порушення або потенціал для значного порушення системи є високим.

Серйозний "червоний" рівень вказує на серйозний рівень загроз від настання кіберінцидентів через зростання хакерських дій, вірусів або іншої зловмисної діяльності, які націлені або компрометують основну інфраструктуру, спричиняють різноманітні перебої надання послуг, різноманітні компрометації систем або ОКІ.

На цьому рівні використовуються вразливості із небезпечним ступенем і поширеним рівнем пошкодження, або порушення чи потенціал для серйозного порушення є високим.

Надзвичайний "чорний" рівень вказує на найвищий рівень загроз від настання кіберінцидентів через зростання хакерських дій, вірусів або іншої зловмисної діяльності, внаслідок яких дуже поширюються перебої і/або значна деструктивна компрометація систем невідомими засобами або послаблюється один чи більше секторів КІ.

На цьому рівні використовуються вразливості із небезпечним ступенем або поширеним рівнем пошкодження чи порушення ОКІ.

5. Заходи етапів реагування на кіберінциденти виконуються для кожного рівня кіберзагрози.

У випадку отримання випереджувальної інформації про підготовку та безпосередню загрозу проведення кібератак проти ОКІІ держави з метою випереджувального реагування, нарощування готовності відповідних сил та засобів рішення про введення необхідних етапів реагування на кіберінциденти приймається НКЦК.

У випадку раптового початку проведення кібератак проти ОКІІ держави рішення про введення відповідних етапів реагування на кіберінциденти приймаються керівництвом суб'єкта забезпечення КБ, силами та засобами якого виявлено факти проведення зазначених протиправних дій в кіберпросторі (кібератак, кіберінцидентів тощо), про що невідкладно інформуються інші СЗК.

Інформація про конкретний ОКІ, щодо якого стався кіберінцидент, є інформацією з обмеженим доступом.

Враховуючи зазначене, є необхідним визначити додаткові завдання окремих суб'єктів взаємодії на кожному етапі реагування на кіберінциденти, а саме.

Під час Етапу 1 (підготовка):

1). ДССЗЗІ України:

координує діяльність інших суб'єктів взаємодії щодо кіберзахисту;

здійснює оцінку стану захищеності державних інформаційних ресурсів в інформаційно-комунікаційних системах та виявляє можливі уразливі місця програмно-апаратних засобів, які використовуються для обробки інформації (місця, використовуючи які зловмисник може порушити цілісність, доступність, конфіденційність інформації або спостережність системи);

надає рекомендації (у тому числі, шляхом розміщення на своєму офіційному веб-сайті), консультативно-методичну і практичну допомогу суб'єктам/операторам критичної інформаційної інфраструктури з питань протидії кіберзагрозам та кіберзахисту, зокрема щодо усунення вразливостей, виявлених за результатами проведення оцінок стану захищеності держав) цих інформаційних ресурсів;

накопичує та проводить аналіз даних про кіберінциденти, а також веде інтерактивну базу даних про кіберінциденти (державний реєстр кіберінцидентів);

інформує інших суб'єктів взаємодії про кіберзагрози;

організує та проводить практично семінари з питань кіберзахисту для суб'єктів взаємодії;

взаємодіє з іноземними та міжнародними організаціями з питань реагування на кіберінциденти, зокрема в рамках участі у Форумі команд реагування на інциденти безпеки FIRST;

розробляє, супроводжує і поширює між основними СЗК модель технічних розвідок іноземних держав, що здійснюють свою діяльність у кіберпросторі.

2). Міністерство оборони України та Генеральний штаб Збройних Сил України:

здійснюють заходи із підготовки держави до відбиття воєнної агресії у кіберпросторі (КО) [11], координують діяльність державних органів та органів місцевого самоврядування щодо підготовки та ведення КО;

отримують від основних СЗК та узагальнюють інформацію щодо ОКІІ воєнної сфери та сфери оборони держави;

проводить інформаційно-аналітичну діяльність та прогнозування розвитку обстановки у воєнній сфері, пов'язану з кіберзагрозами та кіберпростором;

підтримують сили та засоби для дій в кі0ерпросторі в готовності до виконання завдань за призначенням, здійснюють адекватне нарощування їх готовності в залежності від рівня загроз та ступенів реагування на них;

забезпечують несення бойового чергування визначених сил та засобів в інтересах підготовки та ведення КО;

здійснюють підготовку та застосування Збройних Сил України в кіберпросторі щодо виконання ними завдань за призначенням та безпечного використання ними кіберпростору;

здійснюють розвиток необхідних спроможностей Міністерства оборони України, Збройних Сил України для дій в кіберпросторі, підготовки та ведення КО, створення та розвиток відповідних організаційних структур, їх комплектування, підготовку та всебічні: забезпечення;

здійснюють військову співпрацю з НАТО, пов'язану з безпекою кіберпростору та спільним захистом від кіберзагроз, в тому числі й з військовими CERT країн-членів НАТО.

3). Розвідувальні органи України:

здійснюють розвідувальну діяльність щодо загроз національній безпеці України у кіберпросторі, інших подій і обставин, що стосуються сфери КБ;

надають в установленому законодавством порядку основним суб'єктам забезпечення КБ інформацію щодо виявлених в ході здійснення розвідувальної діяльності зовнішніх загроз національній безпеці у кіберпросторі;

подають ДССЗЗІ України встановленим порядком розвідувальну інформацію про технічні розвідки іноземних держав, які діють у кіберпросторі.

4). СБ України:

здійснює відповідно до законодавства контрозвідувальну діяльність із запобігання розвідувально-підривним терористичним та іншим посяганням на КБ України;

інформує основних СЗК про організацію, сили, засоби, методи, тактику розвідувально-підривної діяльності технічних розвідок іноземних держав, міжнародних та іноземних терористичних угруповань, які діють у кіберпросторі, що стали відомими в ході контррозвідувального забезпечення КБ держави;

негласно перевіряє готовність ОКІ до масованих кібератак та кіберінцидентів та інформує ДССЗЗІ України про виявлені у процесі контррозвідувальної діяльності вразливості, що становлять загрозу безпеці ОКІІ;

інформує суб'єктів/операторів критичної інформаційної інфраструктури про розкриті злочини, спрямовані проти безпеки їхніх інформаційних, комунікаційних та інформаційно-комунікаційних систем, умови, що сприяють реалізації кіберзагроз, можливі причини виникнення таких умов та шляхи їхнього усунення.

5). Національна поліція України:

інформує основних СЗК про організацію, сили, засоби, методи, тактику дій злочинних угруповань, що стали відомими в ході оперативно-розшукової діяльності та при обміні інформацією з правоохоронними органами іноземних держав та міжнародних правоохоронних органів (Європол, Інтерпол, тощо);

проводить профілактичні (попереджувальні) заходи із забезпечення КБ ОКІ, а також роз'яснювальну роботу серед всіх верств населення;

повідомляє основних СЗК про виявлені у процесі оперативно-розшукової діяльності вразливості, що становлять загрозу безпеці ОКІІ;

інформує суб'єктів/операторів критичної інформаційної інфраструктури про виявлені у процесі оперативно-розшукової діяльності посягання на безпеку їхніх інформаційних, комунікаційних та інформаційно- комунікаційних систем, умови, що сприяють реалізації кіберзагроз, можливі причини виникнення таких умов та шляхи їхнього усунення.

6). Сили кіберзахисту:

з урахуванням інформації, отриманої від основних СЗК аналізують ризики, впроваджують та вдосконалюють заходи з кіберзахисту;

здійснюють моніторинг кіберзагроз та виявлення кіберінцидентів;

проводять навчання та тренінги фахівців з кіберзахисту, зокрема з питань моніторингу кіберзагроз та виявления кіберінцидентів.

7). Власники та/або керівники ОКІ:

проводять оцінку поточного стану ЗІ та кіберзахисту ОКІІ (прогнозування виникнення нових кіберзагроз, їх врахування в моделі загроз, визначення необхідності її коригування тощо), розраховують ризики КБ;

на підставі аналізу розрахованих ризиків КБ здійснюють практичні заходи щодо забезпечення ЗІ та кіберзахисту ОКІІ з урахуванням інформації, отриманої від основних СЗК та/або суб'єктів кіберзахисту [9];

здійснюють моніторинг, реєстрацію та аудит подій на ОКІІ;

супроводжують та актуалізують еталонні, архівні і резервні копії програмних компонентів, забезпечують зберігання резервних копій даних;

забезпечують виконання персоналом і користувачами вимог, норм, правил, інструкцій з ЗІ відповідно до визначеної політики безпеки;

розробляють плани відновлення сталого функціонування своїх ОКІІ з розрахованим цільовим часом відновлення, у разі порушення його функціонування внаслідок реалізації кібератаки;

надають на запит ДССЗЗІ України необхідну інформацію про реалізовані заходи щодо кіберзахисту ОКІІ.

Під час Етапу 2 (виявлення та аналіз):

1). ДССЗЗІ України:

координує діяльність інших суб'єктів взаємодії щодо вжиття необхідних заходів з кіберзахисту з урахування виявлених кіберзагроз щодо ОКІІ;

забезпечує реагування на кібератаки (кіберінциденти), залучаючи, за необхідності, можливості суб'єктів кіберзахисту;

інформує Національну поліцію та СБ України про об'єкт та джерело походження кібератаки, а суб'єктів/операторів критичної інформаційної інфраструктури щодо таких фактів;

обробляє та накопичує дані про вчинення та/або спроби вчинення кібератак (кіберінцидентів).

2). Міністерство оборони України та Генеральний штаб Збройних Сил України з отриманням інформації про об'єкт га джерело кібератаки на об'єкти воєнної сфери або сфери оборони держави:

здійснюють з основними суб'єктами забезпечення КБ підготовку та проведення заходів щодо КО;

вживають заходів щодо КО (активного кіберзахисту);

інформують ДССЗЗІ України, CERT-UA, СБ України, Національну поліцію України, розвідувальні органи України та інших СЗК про ймовірні об'єкти та джерело кібератаки.

3). СБ України:

виявляє спеціальними методами та засобами кібератаки на ОКІІ, надає їм оперативну та правову оцінку, перевіряє отриману інформацію стосовно їх спрямованості, мотивів, суб'єктів, засобів, методів, тактики, можливих наслідків, умов та чинників, що сприяли їх здійсненню;

повідомляє ДССЗЗІ України первинну інформацію про виявлені кібератаки та кіберінциденти, інформує про результати її оперативної та правової оцінки, пропозиції щодо вжиття невідкладних заходів кіберзахисту, а також інші відомості, необхідні для вжиття зазначених заходів.

4). Національна поліція України:

відповідно до законодавства здійснює заходи щодо розшуку та виявлення осіб, підозрюваних у скоєнні злочину, на підставі інформації про об'єкт та джерело кібератаки, отриманої від інших суб'єктів взаємодії;

інформує суб'єктів взаємодії про виявлені у процесі оперативно-розшукової діяльності про об'єкт та джерело кібератаки на ОКІІ.

5). Сили кіберзахисту:

у разі виявлення кіберінцидентів або фактів здійснення кібератак, негайно інформують щодо таких подій всих суб'єктів взаємодії;

здійснюють блокування джерел кібератак та кіберінцидентів;

інформують ДССЗЗІ України, СБ України та Національну поліцію України про об'єкт та джерело кібератаки для вжиття заходів із запобігання та припинення кіберзлочинів;

здійснюють обробку, накопичення та аналіз даних про спроби та/або факти вчинення кібератак (кіберінцидентів), а також про їх наслідки.

6). Власники та/або керівники ОКІ:

здійснюють невідкладне (протягом однієї години) інформування суб'єктів взаємодії про виявлені кіберінциденти чи спроби та/або факти вчинення кібератак;

негайно втручаються у разі виявлення кібератаки у процес функціонування інформаційно-комунікаційних систем з метою мінімізації наслідків;

зберігають (фіксують) ознаки кібератаки (кіберінцидента), у тому числі на матеріальних носіях інформації.

Під час Етапу 3 (стримування):

1). ДССЗЗІ України:

надає консультативно-методичну допомогу суб'єктам кіберзахисту і суб'єктам/операторам критичної інформаційної інфраструктури з питань реагування на кіберінциденти та заходів, які необхідно вжити для стримування кібератак.

2). Власники та або керівники ОКІ:

підсилюють захист найбільш важливих сервісів, проводять екстрені заходи із забезпечення безпеки внутрішньої мережі, захисту периметра;

забезпечують своєчасне та безперешкодне ознайомлення представників Національної поліції та СБ України з виявленими слідами протиправної діяльності в кіберпросторі для їх оперативного аналізу.

Під час Етапу 4 (усунення):

1). ДССЗЗІ України:

надає консультативно-методичну і практичну допомогу суб'єктам/операторам критичної інформаційної інфраструктури та суб'єктам кіберзахисту, координує їх дії щодо припинення кібератаки або кіберінцидента (за необхідності з виїздом на місце події);

здійснює у разі необхідності практичні заходи, спрямовані на кіберзахист ОКІІ та усунення наслідків кібератак і кіберінцидентів;

вивчає спільно з Національною поліцією та СБ України механізми виявлених кіберінцидентів і кібератак, оцінює негативні наслідки та розробляє шляхи їхньої локалізації;

здійснює взаємодію з суб'єктами кіберзахисту, а також міжнародну взаємодію з командами реагування (CERT, CSIRT) інших країн щодо припинення (блокування) кібератак (кіберінцидентів) та усунення їхніх наслідків.

2). Міністерство оборони України та Генеральний штаб Збройних Сил України:

організовують та здійснюють заходи з кіберзахисту об'єктів воєнної сфери або сфери оборони держави, а також практичні заходи щодо усунення наслідків реалізації кібератак і кіберінцидентів;

надають консультативно-методичну та практичну допомогу підрозділам воєнної сфери та сфери оборони щодо припинення та усунення наслідків кібератаки або кіберінцидента (за необхідності з виїздом на місце події);

забезпечують безпосередню взаємодію з військовими CERT країн-членів НАТО щодо припинення кібератак (кіберінцидентів).

3). СБ України:

вивчає спільно з ДССЗЗІ України та Національною поліцією України механізми виявлених кіберінцидентів і кібератак, долучається до оцінки негативних наслідків та розробки шляхів їх локалізації;

інформує інших суб'єктів взаємодії про виявлені причини виникнення (здійснення) кіберінцидентів і кібератак, умови, що цьому сприяли, та шляхи їхнього усунення.

4). Сили кіберзахисту:

здійснюють, з урахуванням інформації отриманої від ДССЗЗІ України та СБ України, практичні заходи з кіберзахисту ОКІІ та усувають наслідки кібератаки або кіберінцидента;

надають консультативно-методичну та практичну допомогу суб'єктам/операторам критичної інформаційної інфраструктури щодо припинення та усунення наслідків кібератаки або кіберінцидента (за необхідності з виїздом на місце події);

здійснюють взаємодію з ДССЗЗІ України, а також міжнародну взаємодію з командами реагування інших країн (CERT, CSIRT, MCSIRT) щодо припинення (блокування) кібератак (кіберінцидентів) та усунення їхніх наслідків.

5). Власники та/або керівники ОКІ:

усувають наслідки кібератак (кіберінцидентів) з урахуванням інформації, отриманої від інших суб'єктів взаємодії.

Під час Етапу 5 (відновлення):

1). ДССЗЗІ України:

координує діяльність інших суб'єктів взаємодії щодо кіберзахисту під час відновлення сталого функціонування ОКІІ.

2). СБ України:

проводить заходи з документування фактичних даних про кібератаки, які могли призвести або призвели до вчинення кримінальних правопорушень, криміналістичне дослідження матеріалів, пов'язаних з кібератаками чи кіберінцидентами, здійснює оперативний розшук осіб, причетних до їх підготовки або скоєння;

негласно оцінює стан готовності ОКІ до реагування на кібератаки та кіберінциденти.

3). Національна поліція України:

відповідно до законодавства здійснює заходи щодо встановлення осіб, підозрюваних у скоєнні злочину, та притягнення їх до відповідальності;

вивчає спільно з ДССЗЗІ України та СБ України механізми виявлених кіберінцидентів і кібератак, долучається до оцінки негативних наслідків та розробки шляхів їхньої локалізації;

проводить аналіз подій, спрямований на встановлення причин та передумов виявлених кіберінцидентів і кібератак.

4). Власники та/або керівники ОКІ:

здійснюють власними силами відновлення сталого функціонування інформаційно-комунікаційних систем, виведених з ладу внаслідок кібератак (кіберінцидентів) після нейтралізації загроз, узгоджуючи такі дії з ДССЗЗІ України та/або суб'єктами кіберзахисту (відповідно до підпорядкованості);

забезпечують, у разі необхідності, фізичний доступ представників ДССЗЗІ України, суб'єктів кіберзахисту (відповідно до підпорядкованості) до інформаційно-комунікаційних систем для виконання заходів щодо блокування та локалізації негативних наслідків кібератак (кіберінцидентів) та відновлення сталого функціонування цих систем.

Під час Етапу 6 (заходи після інциденту):

1). ДССЗЗІ України:

здійснює аналіз даних про спроби та/або факти вчинення кібератак (кіберінцидентів), а також про їхні наслідки;

проводить актуалізацію державного реєстру кіберінцидентів з урахуванням нових даних про спроби та/або факти вчинення кібератак (кіберінцидентів), а також про їхні наслідки;

здійснює обмін інформацією з субє'ктами кіберзахисту, а також з командами реагування (CERT, CSIRT) інших країн щодо виявлених кібератак та кіберінцидентів та проведених заходів попередження реалізації кіберзагроз;

готує та надає суб'єктам кіберзахисту та суб'єктам/операторам критичної інформаційної інфраструктури практичних рекомендацій за результатами аналізу даних про спроби та/або факти вчинення кібератак (кіберінцидентів), а також про їхні наслідки.

2). Міністерство оборони України та Генеральний штаб Збройних Сил України:

встановленим порядком отримують та узагальнюють інформацію щодо результатів підготовки та ведення КО;

здійснюють підготовку та надання рекомендацій щодо попередження реалізації кіберзагроз у воєнній сфері та сфері оборони;

забезпечують взаємодію з військовими CERT країн-членів НАТО та з виконавчими підрозділами СЗК з питань ЗІ, кіберзахисту, КБ та КО.

3). СБ України:

надає ДССЗЗІ України підготовлену на основі оперативних матеріалів узагальнену інформацію щодо фактичної готовності ОКІІ до можливих кібератак (кіберінцидентів), а також обґрунтовані пропозиції щодо її поліпшення;

надає ДССЗЗІ України прогностичну інформацію щодо можливих в подальшому кібератак та кіберінцидентів, а також рекомендації щодо заходів кіберзахисту.

4). Національна поліція України:

інформує громадян про заходи щодо забезпечення безпеки в кіберпросторі;

надає рекомендації суб'єктам кіберзахисту та суб'єктам/операторам критичної інформаційної інфраструктури, громадянам стосовно запобігання кіберзлочинам.

5). Суб'єкти кіберзахисту:

аналізують та проводять експертну оцінку даних про спроби та/або факти вчинення кібератак (кіберінцидентів), способи реалізації кібератак і кіберінцидентів, розробляють заходи з протидії кібератакам і кіберінцидентам;

ведуть власні бази даних кіберінцидентів, забезпечують передачу відповідної інформації до загальної інтерактивної бази даних про кіберінциденти (державного реєстру кіберінцидентів);

здійснюють взаємодію з ДССЗЗІ України та командами реагування (CERT, CSIRT) інших країн з питань попередження кіберзагроз (кіберінцидентів);

здійснюють підготовку та надання суб'єктам/операторам критичної інформаційної інфраструктури практичних рекомендацій щодо попередження кібератак (кіберінцидентів).

6). Власники та/або керівники ОКІ: здійснюють збір, узагальнення та аналіз інформації про кібератаки (кіберінциденти) та подають її до ДССЗЗІ України та суб'єктам кіберзахисту (відповідно до підпорядкованості);

розраховують ризики КБ, на підставі розрахунків вдосконалюють політики безпеки та розробляють нові заходи з протидії кібератакам і кіберінцидентам.

6. Методичні рекомендації щодо змісту заходів за етапами реагування на кіберінциденти, типи (таксономію) кіберінцидентів, загальні правила обміну інформацією про кіберінциденти затверджує Адміністрація ДССЗЗІ України.

Висновки з даного дослідження і перспективи подальших досліджень у даному напрямку

Отже за результатами проведеної роботи в даній публікації були досліджені передумови і особливості формування законодавства України у сфері КБ, визначені проблеми та перспективи його подальшого розвитку з точки зору оцінки наявних небезпек та загроз. Визначені напрями адаптації чинного законодавства про КБ до стандартів ЄС у межах реалізації положень Угоди про асоціацію між Україною та ЄС та представлено рекомендації, щодо попередження, виявлення та порядку реагування на існуючі загрози національному сегменту кіберпростору держави.

Також були визначені найбільш перспективні напрями розвитку національної системи кіберзахисту, а саме вдосконалення правової основи кіберзахисту ОКІ; впровадження системи незалежного аудиту інформаційної безпеки на ОКІ; створення галузевих центрів реагування на кіберінциденти; розвиток міжнародного співробітництва у сфері забезпечення КБ в яких держава повинна відігравати сервісну роль.

Список літератури

1. Про внесення змін до Указів Президента України від 27 січня 2015 року № 37 та від 7 червня 2016 року № 242: Указ Президента України №27/2020 від 28 січ. 2020 р. URL: https://www.president.gov.ua/documents/272020-32041.

2. Президент увів у дію рішення РНБО про захист від кібератак, 2017 р. URL: https://www.ukrinform.ua/rubric-polytics/2296115-prezident-uviv-u-diu-risenna-rnbo-pro-zahist-vid-kiberatak.html.
3. Порядок взаємодії суб'єктів забезпечення кібербезпеки під час реагування на кіберінциденти/кібератаки одноголосно затверджено на засіданні НКЦК, 28 вер. 2022 р. URL: https://www.rnbo.gov.ua/ua/Diialnist/5765.html.
4. Про критичну інфраструктуру: закон України док. 1882-IX, від 16 листоп. 2021 р. URL: https://zakon.rada.gov.ua/laws/show/1882-20#Text.
5. Про затвердження плану заходів на 2017 рік з реалізації Стратегії кібербезпеки України: Розпорядження Кабінету міністрів України від 10 березня 2017 р. № 155-р. URL: https://zakon.rada.gov.ua/laws/show/155-2017-%D1%80#Text.
6. Грабовий А. М. Закон про кібербезпеку та стратегія кібербезпеки України. Онлайн-видання Юрист&Закон. 2022. №28. URL: https://uz.ligazakon.ua/ua/magazine_article/EA010553.
7. Онищук І.І. Особливості бюджетного процесу в умовах воєнного стану. Пресцентр ініціативи "Децентралізація". 2022. URL: https://decentralization.gov.ua/news/14654.
8. Про затвердження Порядку проведення огляду стану кіберзахисту критичної інформаційної інфраструктури, державних інформаційних ресурсів та інформації, вимога щодо захисту якої встановлена законом: Постанова Кабінету міністрів України від 11 листопада 2020 р. № 1176. URL: https://zakon.rada.gov.ua/laws/show/1176-2020-%D0%BF#Text.
9. Про рішення Ради національної безпеки і оборони України від 27 січня 2016 року "Про Стратегію кібербезпеки України": Указ Президента України №96/2016р. ред. від 28 серпня 2021. URL: https://www.rnbo.gov.ua/ua/Ukazy/417.html.
10. Про затвердження Положення про організацію кіберзахисту в банківській системі України та внесення змін до Положення про визначення об'єктів критичної інфраструктури в банківській системі України: Постанова правління Національного банку України від 12 серпня 2022 р. № 178. URL: https://zakon.rada.gov.ua/laws/show/v0178500-22#Text.
11. Тютюнік В., Горовенко В. Всеохоплююча оборона України: стан, проблеми та заходи щодо зміцнення кібероборони держави і створення кібервійськ. Оборонно-промисловий кур'єр. 01 листопада 2021 р. URL: https://opk.com.ua/%D0%B2%D1%81%D0%B5%D0%BE%D1%85%D0%BE%D0%BF%D0%BB%D1%8E%D1%8E%D1%87%D0%B0-%D0%BE%D0%B1%D0%BE%D1%80%D0%BE%D0%BD%D0%B0-%D1%83%D0%BA%D1%80%D0%B0%D1%97%D0%BD%D0%B8-%D1%81%D1%82%D0%B0%D0%BD-%D0%BF%D1%80/.
12. Онлайн-шахрайство з використанням тематики "допомоги від Червоного Хреста" (CERT-UA#5063): CERT-UA Державної служби спеціального зв'язку та захисту інформації України. 27 липня 2022 р. URL: https://cert.gov.ua/article/987552.

Received (Надійшла) 20.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Protocol of general actions of cyber security subjects during response to cyber incidents, as well as in the elimination of their consequences

Y. Zhyvylo, I. Romashko

Abstract. Cyberspace, along with other physical spaces, is recognized as one of the theaters of war. The tendency to create cyber troops is gaining momentum (Roadmap for the creation of Cyber Troops of the Armed Forces of Ukraine - Order of the General Staff of the Armed Forces of Ukraine, dated April 22, 2022 No. 48), whose tasks include not only ensuring the protection of the state's critical information infrastructure from cyber attacks, but also conducting preventive offensive (carrying out cyber operations) in cyberspace, including the disabling of critical infrastructure facilities of the enemy by destroying the information systems that manage such facilities. An increase in the intensity of interstate confrontation and reconnaissance and subversive activities in cyberspace is predicted. The circle of states that are trying to form their own cyber intelligence, to master modern technologies of reconnaissance and explosive activities in cyberspace is expanding, and they are strengthening state control over national segments of the Internet. Taking into account the experience of conducting hostilities during the introduction of the legal regime of martial law and given the uncertainty of subjects and objects, their functions and tasks for actions in certain areas, including in the field of cybersecurity, in peacetime led to adversity and inconsistency of these actions by the subjects of support state cybersecurity. And given that with the introduction of the legal regime of martial law, certain entities change their location, move information assets and equipment to new locations using cloud services, which greatly complicates the process of harmonization and coordination of actions to respond to cyber incidents, as well as elimination of consequences. This leads to a forced redistribution of tasks and functions for the implementation of cyber defense measures at various facilities. Under these conditions, new subjects of cyber defense are created on a permanent or temporary basis, which requires time for them to acquire the abilities to perform their intended tasks. In such a situation, Ukraine should be able to ensure its socio-economic development in the digital world, which requires the acquisition of the ability to effectively deter destructive actions in cyberspace, a sustainable response to threats in cyberspace, the achievement of cyber resilience at all levels, and the interaction of the components of the security and defense sectors to ensure cybersecurity within the cyber defense of the state. So, based on the need for a scientific justification of the institutional framework, it is necessary to clearly define: "a list of subjects for ensuring cybersecurity for the implementation of the actions established by this Protocol", both in peacetime and under the legal regime of martial law; the above subjects, their role and place, the list and procedure for responding to cyber incidents and eliminating their consequences, both in peacetime and under the legal regime of martial law. At the same time, the scientific novelty of the expected results lies in the theoretical substantiation and provision of practical recommendations for improving the mechanisms for managing and interacting with the components of the security and defense sector when planning the state's preparation for cyber defense, taking measures to neutralize and actively counter cyber threats in the national segment of the state's cyberspace.

Keywords: cybersecurity subjects, cyberspace, cyber defense, active cyber actions, destructive cyber attacks, critical information infrastructure.

УДК 004.9 **doi: 10.26906/SUNZ.2024.1.077**

Д. Я. Зайцев, Т. В. Філімончук, А. С. Гук, Г. В. Майстренко

Харківський національний університет радіоелектроніки, Харків, Україна

ОГЛЯД ЗАСОБІВ ЕФЕКТИВНОЇ СЕГМЕНТАЦІЇ ЗОБРАЖЕНЬ З ВИКОРИСТАННЯМ МЕТОДІВ КЛАСТЕРИЗАЦІЇ ДАНИХ

Анотація. Актуальність. Протягом багатьох десятиліть люди мріяли створити машини, які могли б зрівнятись по характеристикам з людським інтелектом та мислити й діяти як люди. Однією з найцікавіших ідей було дати комп'ютерам можливість «бачити» та інтерпретувати навколишній світ у зручний для розуміння комп'ютером. Завдяки прогресу в галузі штучного інтелекту, інноваціям в області deep learning та нейронних мережах ця сфера змогла зробити значний стрибок за останні роки та перевершити людей у деяких задачах, пов'язаних із виявленням та маркуванням об'єктів. **Метою даної роботи** є розгляд та порівняльний аналіз існуючих засобів сегментації зображень, зокрема, з використанням методів кластеризації. **Об'єктом дослідження** виступають алгоритми сегментації зображень, які є ключовими у сфері комп'ютерного зору для розпізнавання та аналізу об'єктів. **Предметом дослідження** є конкретне використання методів кластеризації в цих алгоритмах, їх ефективність та точність. **Результати.** У даній роботі проведено розгляд та порівняльний аналіз основних методів кластеризації даних. Виявлено 3 найпоширеніші та найперспективніші для покращення алгоритми. Для кожного з них описано принцип роботи, виділено їх головні переваги та недоліки. **Висновок.** Серед розглянутих методів немає найкращого універсального вибору, оскільки ефективність кожного з них залежить від конкретної задачі. Описані алгоритми планують бути використані для їх подальших досліджень та модифікацій.

Ключові слова: кластеризація, сегментація, комп'ютерний зір, обробка зображень, штучний зір.

Вступ

Постановка проблеми. Сучасний технологічний прогрес вражає. Нові досягнення у сферах штучного інтелекту, комп'ютерного зору, Data Mining, активність інтернет-пошуків, постійно зростаюча кількість цифрових зображень та інших медіаматеріалів зародили проблему організації та структуризації цих даних. Окрім того, з часом постала необхідність зручного оперування та аналізу такими даними. Останнім часом активно почала розвиватись сфера розпізнавання зображень. Це сприяло створенню автомобілів, які здатні самокеруватись без участі водія, реалізації систем спостереження, які б могли за лічені секунди розпізнати кожну особу в полі бачення однієї або цілої мережі відеокамер, полегшенню роботи в області рентгенографії алгоритмами, які можуть аналізувати сотні знімків за раз.

Кластеризація вважається одним з найцікавіших підходів до задачі пошуку схожих даних і об'єднання їх у групи. Вона передбачає розбиття набору даних на кілька груп таким чином, що подібність усередині групи є більшою ніж серед усіх груп.

Аналіз останніх досліджень і публікацій. У статті [1] було проведено порівняльний аналіз методів сегментації зображень: розглянуто актуальні типові методи сегментації зображень та аналіз кожного з них; окреслено поняття сегментації, виділено її мету та сферу використання. Також, згідно статті, сегментацію умовно можна поділити на сегментацію статичних та динамічних зображень. В першому випадку доводиться мати справу з окремими зображенням (картинками), а у другому – з відео потоком даних.

Як одну з можливих опцій для сегментації, розглянуто методи, які засновані на використанні порогів [2]. Також у публікації було порівняно ефективність алгоритмів із використанням гістограми з іншими. Окрім того, серед розглянутих методів також є згадка про алгоритми, які базуються на кластеризації, зокрема, алгоритмом K-Means, який є одним з найпопулярніших через задачі кластеризації.

Автори дійшли висновку що існування значної кількості методів сегментації зображень вимагає їх дослідження та формування рекомендацій щодо їх використання в конкретних задачах розпізнавання зображень в різних предметних областях. А також створення нових методів сегментації, які не будуть мати недоліки розглянутих методів, насамперед мова йде про методи адаптивної сегментації.

У статті [3] наведено огляд сучасних методів комп'ютерної або автоматизованої сегментації анатомічних медичних зображень. Коротко описано конкуруючі методи та їх застосування. Також увага зосереджується на представленні актуальних застосувань сегментації зображень в медичній візуалізації та різноманітних проблемах галузі, які мають бути вирішеними.

Незважаючи на те, що автори посилаються здебільшого на найбільш часто використовувані рентгенологічні методи візуалізації анатомії, більшість описаних концепцій також можна застосувати до інших методів сегментації.

Стаття [4] містить комплексний огляд різних технік сегментації зображень. Автори визначають мету процесу сегментації як самостійне розкладання зображення на окремі регіони. Також у статті наведено кілька прикладів застосування даної техніки у сьогоденні, так, згідно думки авторів, сегментація зображення стала дуже важливою задачею в сьогоднішньому сценарії розвитку технологій, у сучасному світі комп'ютерне бачення стало міждисциплінарною областю і його застосування можна знайти в будь-якій сфері.

Що до питаня сегментації зображень, також доречно було б розглянути інші існуючі та конкуруючі методи, окрім кластеризації. Частково, це було зроблено у статті [1], яка саме розглядала та порівнювала існуючі рішення для сегментації зображень, але варто більше звернути уваги на один з них, а саме на генетичні алгоритми.

У статті [5] автори використовують генетичні алгоритми для розв'язання проблеми сегментації зображень, яка є вирішальним етапом у процесі оброблення та аналізу зображень. Згідно статті, сегментація зображення – це процес розбиття одного зображення на множину сегментів, де сегменти вже більш репрезентативні та зручніші для дослідження, як деталі можна використовувати окремі поверхні або предмети. Процес сегментації зображень застосовують для визначення об'єктів та їхніх меж. Сенс використання генетичних алгоритмів полягає у тому, що кожен піксель групується в інші пікселі за допомогою функції відстані на основі як локальних, так і глобальних уже обчислених сегментів. Майже кожен алгоритм сегментації зображень містить параметри, які використовують для управління результатами сегментації; генетична система може динамічно змінювати параметри для досягнення найкращих показників.

Як і в послідовності зображень, для оптимізації декількох параметрів автори застосовували багатоцільові генетичні алгоритми, за допомогою яких можна знайти різноманітну колекцію рішень із більшою кількістю змінних.

Алгоритми кластеризації поширені набагато більше аніж суто на вирішення проблем сегментації. В статі [6] представлена загальна характеристика процедури кластерного аналізу. Наведено огляд існуючих підходів до вирішення задачі кластеризації та математичних методів кластерного аналізу даних. Описано етапи процесу кластеризації, розглянуто питання вибору міри відстані та ваг для класифікуючих властивостей об'єктів. Проведено класифікацію та аналіз існуючих алгоритмів кластерного аналізу, розглянуто переваги та недоліки цих алгоритмів. Обґрунтовано доцільність використання карт Кохонена в методиках кластеризації з метою дослідження наявності чи відсутності кластерної структури в даних, числа кластерів, законів сумісного розподілу ознак, залежностей тощо. Надано порівняльну таблицю алгоритмів та зроблено висновок щодо необхідності подальшого розроблення простих в реалізації алгоритмів, які потребують мінімальної кількості початкових параметрів, дозволяють проводити багатоваріантний аналіз та дають задовільні результати.

Коли мова йде про взаємозв'язок кластеризації та зображень, не завжди мається на увазі сегментація. У статті [7] запропоновано метод кластеризації зображень для їх подальшого адаптивного стискання. Проведено дослідження ефективності стискання зображень шляхом використання кластеризації та стискання кластерів на основі перетворення Карунена-Лоєва. Обраний метод кластеризації є підґрунтям для подальшого стиснення інформації на основі перетворення Карунена-Лоєва, оскільки він виключає можливість перетину кластерів.

Метою цієї роботи є розгляд та порівняльний аналіз існуючих засобів сегментації зображень, зокрема, з використанням методів кластеризації.

Основна частина

Сегментація зображення – це процес поділу цифрового зображення на кілька сегментів (об'єктів). Мета сегментації – змінити подання зображення на щось більш значуще та легше для аналізу.

Сегментація зображення – це важливий крок в обробці зображення, і вона стає майже обов'язковою, якщо є задача проаналізувати, що знаходиться всередині зображення.

Наприклад, якщо необхідно визначити, чи є стілець або людина всередині зображення приміщення, то в цьому випадку може знадобитися сегментація зображення, щоб розділити об'єкти та проаналізувати кожен об'єкт окремо, щоб перевірити, що це таке. Сегментація зображення зазвичай виконує функцію попередньої обробки перед розпізнаванням образів, виділенням ознак та стисненням зображення.

Загалом, кластеризація – це ніщо інше, як групування наданих даних відповідно до їх подібності та отримання різних кластерів у кінці. Відповідно до методу кластеризації, який використовується, спосіб групування даних змінюється. Розглянемо 2 різні типи сегментації зображень, які найчастіше використовуються: кластеризація з розділенням (Partitioning Clustering) та нечітка (Fuzzy Clustering) кластеризація.

Методи кластеризації з розділенням ділять дані на k груп, де k – це деяке число, яке визначається завчасно користувачем. Одним з найпопулярніших представників такого різновиду методів кластеризації є метод K-Means.

Нечітка кластеризація є жорстким типом кластеризації, тоді як кластеризація з розділенням називається м'якою. Причина цього полягає в тому, що в кластеризації з розділенням одна точка даних може мати лише один кластер. У нечіткій кластеризації з вихідними даними існує деяка ймовірність для кожної точки даних, і кожна з точок може належати будь-якому кластеру на цьому рівні ймовірності. Найпоширеніший метод який демонструє нечітку (fuzzy) кластеризацію – це метод C-Means. Цікава назва, бо при схожості назви з вище згаданим алгоритмом K-Means, даний алгоритм базується на іншому алоритмі та відноситься до іншої групи методів кластеризації.

Також не зайвим буде згадати інші 2 менш поширені типи кластеризації, а саме гірну (mountain clustering) та субтрактивну (subtractive clustering).

Mountain clustering обчислює функцію щільності у кожній можливій позиції у просторі даних і серед них знаходить позицію з найбільшою щільністю в якості вершини або центру першого кластера. Потім алгоритм знищує першу вершину та шукає другу. Таким чином алгоритм повторюється допоки

необхідна кількість вершин, які є кластерами, не буде знайдена.

Subtractive clustering дещо схожа за принципом дії до попередньої, за виключенням того, що вона не рахує функцію щільності та не намагається знайти вершину в кожній можливій точці, а шукає її лише серед точок з даними.

Таким чином, кількість обчислень суттєво зменшується. Але попри це, у порівнянні з алгоритмами K-Means та C-Means, гірний та субстрактивний методи майже не використовуються.

Кластеризація методом K-Means Кластеризація методом K-Means – це алгоритм кластеризації, який відноситься до групи алгоритмів що не потребують навчання, він групує набір даних без міток у різні кластери. Тут k визначає кількість заздалегідь визначених кластерів, які необхідно створити в процесі, наприклад, якщо k=2, буде два кластери, а для k=3 буде три кластери і так далі.

Це дозволяє кластеризувати дані в різні групи та по суті є зручним способом самостійного виявлення категорій груп у немаркованому наборі даних без необхідності навчання.

Даний алгоритм базується на основі центроїда, де кожен кластер пов’язаний із центроїдом. Основною метою цього алгоритму є мінімізація суми відстаней між точкою даних і відповідними кластерами.

Алгоритм приймає набір даних без міток як вхідні дані, ділить набір даних на k кластерів і повторює процес, доки не знайде найкращі кластери. Значення k має бути заздалегідь визначене в цьому алгоритмі. Метод кластеризації K-Means виконує дві задачі:

- визначає найкраще значення для K центральних точок або центроїдів за допомогою ітераційного процесу;

- призначає кожну точку даних найближчому k-центру. Ті точки даних, які знаходяться поблизу певного k-центру, створюють кластер d.

Таким чином, кожен кластер має точки даних з деякими спільними рисами, і він знаходиться далеко від інших кластерів.

На рис. 1 демонструється робота алгоритму кластеризації K-Means.

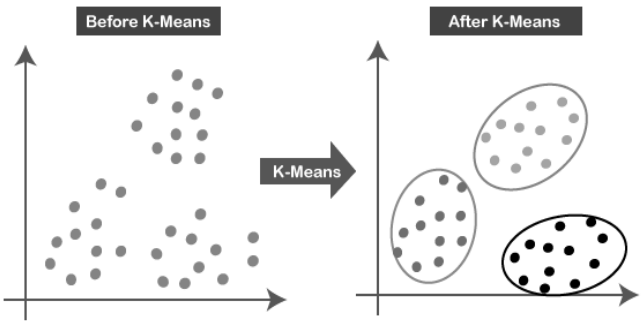


Рис. 1. Результат кластеризації за допомогою методу K-Means

Алгоритм K-Means працює наступним чином:

- на першому кроці визначається число k, щоб встановити кількість майбутніх кластерів;

- на другому кроці здійснюється вибір випадкових k точок або центроїдів (можуть бути відмінні від вхідного набору точки);

- на третьому кроці здійснюється призначення кожної точки даних найближчому центроїду, що сформує попередньо визначені k кластерів;

- на четвертому кроці обчислюється дисперсія та розміщується новий центроїд для кожного кластеру;

- далі виконується повторення третього кроку, тобто перепризначення кожної точки даних новому найближчому центроїду кожного кластера. Якщо відбувається будь-яке перепризначення, здійснюється повторення четвертого кроку. Інакше, якщо перепризначень більше не відбувається, то модель сформована та кластери призначені.

Продуктивність алгоритму кластеризації методом K-Means дуже залежить від кластерів, які він формує. Підбір оптимальної кількості кластерів може стати проблемним. Існують різні способи визначення оптимальної кількості, тобто значення коефіцієнту K, але найбільш раціональний з них тільки один.

Наведений метод має назву Elbow Method або метод ліктя. Він є одним із найпопулярніших способів знайти оптимальну кількість кластерів і для цього використовує концепцію значення WCSS (within-cluster sum of squares). WCSS – це сума квадратів у кластері, що є кількістю загальних варіацій в межах кластера.

Щоб виміряти відстань між точками даних і центроїдом, можна використовувати будь-який відомий користувачеві метод, наприклад евклідову або манхеттенську відстань.

Для пошуку оптимального значення кластерів, метод ліктя виконує наступні дії:

- здійснює кластеризацію K-Means на заданому наборі даних для різних значень K (діапазони від 1 до 10);

- для кожного значення K обчислюється значення WCSS;

- далі будується крива між обчисленими значеннями WCSS та кількістю кластерів K. Гостра точка вигину або точка графіка, що виглядає як плече, вважається найкращим значенням K (рис. 2).

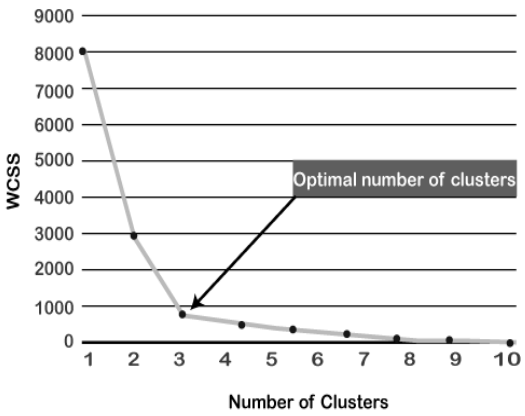


Рис. 2. Графік залежності значення WCSS від кількості кластерів

Fuzzy C-Means кластеризація. Існують алгоритми кластеризації з використанням нечіткої логіки. Тобто, логіки, при якій немає однозначної належності точки конкретному кластеру. Принципи нечіткої логіки можна використовувати для кластеризації багатовимірних даних, призначаючи кожній точці членство в кожному центрі кластера від 0 до 100 відсотків. Це може бути більш ефективно у порівнянні з традиційною кластеризацією з жорстким порогом, де кожній точці призначається чітка, точна мітка.

Алгоритм кластеризації методом C-Means працює шляхом призначення членства кожній точці даних, що відповідає кожному центру кластера, на основі відстані між центром кластера та точкою даних. Чим більше даних знаходиться ближче до центру кластера, тим більше їх приналежність до конкретного центру кластера. Зрозуміло, що сума належності кожної точки даних повинна дорівнювати одиниці.

C-Means – цей метод кластеризації, який дозволяє створювати нечіткі розділи з даних. Алгоритм залежить від параметра m, який відповідає ступеню нечіткості рішення.

Великі значення m призведуть до розмивання меж кластерів, і таким чином всі елементи матимуть тенденцію належати до всіх кластерів. Тобто рішення оптимізаційної задачі залежать від параметра m: різні вибори m зазвичай призведуть до різних розділів.

Нечітка кластеризація методом C-Means або Fuzzy C-Means (FCM) спирається на основну ідею Hard C-Means (HCM) кластеризації з тією різницею, що в FCM кожна точка даних належить кластеру на деякому рівні приналежності, а в HCM кожна точка даних або належить до певного кластеру, або ні. Таким чином FCM використовує нечітке розділення, тобто кожна точка даних може належати до кількох груп зі своїм ступенем приналежності, визначеним оцінкою між 0 та 1.

Однак, FCM все ще використовує доволі затратну функцію, яка має бути мінімізована під час спроби розділення набору даних.

Алгоритм працює ітеративно через попередні дві умови до моменту поки покращення не перестануть спостерігатись.

Порівняємо більш детально алгоритми FCM та K-Means, щоб отримати чітке уявлення про те, де підходить описуваний алгоритм C-Means.

Перше порівняння стосується віднесення до кластера: у нечіткій кластеризації кожна точка має ймовірність належати до кожного кластера, а не повністю належати лише одному кластеру, як це має місце в традиційному алгоритму k-середніх. У FCM кожна точка має вагове значення, пов'язане з певним кластером, тому точка не знаходиться «в кластері» повністю, оскільки вона має слабкий або сильний зв'язок із кластером, який визначається зворотною відстанню до центру кластера.

Друге порівняння стосується швидкості: засоби FCM, як правило, працюватимуть повільніше, ніж засоби k-середніх, оскільки вони насправді виконують більше роботи. Кожна точка оцінюється з кожним кластером, і в кожній оцінці бере участь більше операцій. Метод K-Means потребує обчислення відстані, тоді як нечіткий FCM потребує повного оберненого зважування відстані.

Ієрархічна кластеризація. Ієрархічна кластеризація є однією з найпопулярніших і простих для розуміння методів кластеризації. Ця техніка кластеризації поділяється на два види: агломеративний та розділовий.

У техніці агломеративної ієрархічної кластеризації спочатку кожна точка даних розглядається як окремий кластер. На кожній ітерації подібні кластери зливаються з іншими кластерами, поки не буде сформований один або K кластерів.

Базовий алгоритм Agglomerative має наступні кроки:

- обчислення матриці близькості;
- визначення кожної точки даних як кластера;
- поєднання двох найближчих за відстанню кластерів між собою;
- повторення попередніх кроків допоки необхідна кількість кластерів не буде сформована.

Візуально алгоритм ієрархічної кластеризації демонструють у вигляді дендрограми (рис. 3).

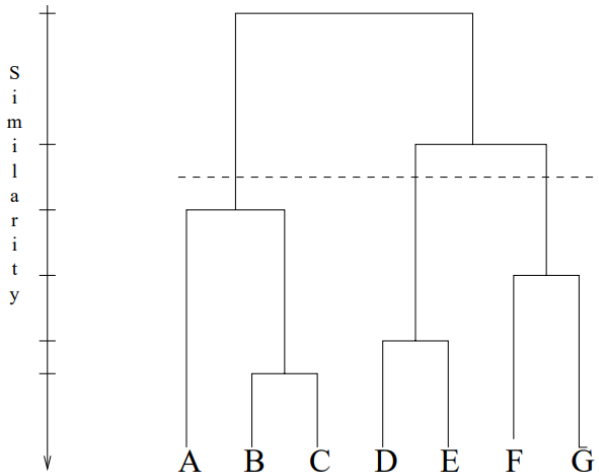


Рис. 3. Дендрограма роботи алгоритму ієрархічнох кластеризації

Наступним методом для розгляду э розділовий алгоритм ієрархічної кластеризації. Можна сказати, що розділова ієрархічна кластеризація є прямою протилежністю агломеративному підходу ієрархічної кластеризації. У розділовій ієрархічній кластеризації всі точки даних розглядаються як один кластер, і на кожній ітерації точки, які найбільш не схожі, відокремлюються від кластера. Кожна відокремлена точка даних розглядається як окремий кластер. Зрештою, залишиться n кластерів, кількість яких задаються на початку. Таким чином, оскільки окремі кластери діляться на n кластерів, це називається розділовою ієрархічною кластеризацією.

Висновки

В результаті проведених авторами досліджень було розглянуто кілька методів кластеризації даних,

серед яких виділено 3 найпоширеніші методи: ієрархічний метод, метод K-Means та C-Means. Відібрані методи вирішують проблему категоризації даних шляхом поділу набору даних на ряд кластерів на основі певної міри подібності, так, що подібність у кожному кластері більша, ніж серед кластерів.

Розглянуті алгоритми кластеризації можна логічним чином поділити на дві групи: ті, які потребують початкового значення кластерів, та ті, яким воно не обов'язкове. Методи K-Means та C-Means можна віднести до алгоритмів з обов'язковим початковим значенням кількості кластерів, а ієрархічні алгоритми, які теж мають деяке число підвидів, до типу алгоритмів, яким значення кількості кластерів не потрібне.

Методи K-Means та C-Means також мають свої відмінності у принципі роботи, ефективності, результативності та призначенні. Таким чином, метод K-Means має більшу швидкодію, але він жорстко поділяє точки даних на кластери, що не завжди може бути зручним. Метод C-Means при цьому може відносити точки до кластерів лише з якимось значенням приналежності, що дозволяє будувати нечіткі, але при цьому більш реалістичні межі між вибірками даних, але при цьому швидкість його роботи значно менша.

Щодо ієрархічної кластеризації, то складність та швидкість роботи ще менша за попередні розглянуті методи, але при цьому початкове значення кластерів не є обов'язковим, залежно від конкретного підвиду ієрархічної кластеризації.

Таким чином, можна прийти до висновку, що однозначного фавориту серед розглянутих методів немає, бо не існує єдиного критерію якості кластеризації.

Варто зазначити, що вибір методу значною мірою залежить від специфіки даних та поставлених задач.

Кожен підхід може давати різні результати, і кожен з них найкраще проявить свою ефективність тільки для обмеженого набору задач.

Також, неієрархічні методи часто базуються на початковому значенні кластерів, що не завжди є можливим або потрібним.

Під час вибору методу, який буде використовуватись варто враховувати далеку від лінійної складність, особливо це стосується ієрархічних типів методів кластеризації.

Додатково слід враховувати потребу в інтерпретації результатів, оскільки в залежності від використаного алгоритму, отримані набори кластерів можуть різнитися між собою.

СПИСОК ЛІТЕРАТУРИ

1. Ониськів П.А., Литвиненко Я.В. (2019), "Аналіз методів сегментації зображень", *Матеріали IV Міжнародної науково-технічної конференції „Теоретичні та прикладні аспекти радіотехніки, приладобудування і комп'ютерних технологій" присвячена 80-ти річчю з дня народження професора Я.І. Проця*, С. 48-49.
2. Grady L. (2006), "Random walks of image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Pp. 1768-1783.
3. Dzung L.P., Chenyang Xu., Jerry L.P. (2000), "Current Methods in Medical Image Segmentation", *Annual Review of Biomedical Engineering Vol. 2*, Pp. 315-337.
4. Tara S., Reddy B., Ramesh G., Sandeep K. (2014), "Various Image Segmentation Methods Based On Partial Differential Equation-A Survey", *International Conference on Computer & Communication Technologies Vol. 3*, Pp. 183-186.
5. Гороховський С.С., Мороз А.В. (2021), "Сегментація зображень із використанням генетичних алгоритмів", С. 52-55.
6. Волосюк Ю.В. (2014), "Аналіз алгоритмів кластеризації для задач інтелектуального аналізу даних", *Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка*, С. 112-119
7. Скаковська А.М., Радивоненко О.С., Шалда К.В. (2012), "Кластеризація зображень для їх компресії на основі компонентного аналізу", *Вісник Сумського державного університету*, С. 32-36.
8. Jain A.K., Murty M.N., Flynn P.J. (1999), "Data clustering: a review", *ACM Computing Surveys*, Vol. 31, No. 3, September 1999, pp. 264-323.

Received (Надійшла) 22.12.2023
(Accepted for publication) Прийнята до друку 24.01.2024

Review of effective image segmentation using data clustering methods

Dmytro Zaitsev, Tetiana Filimonchuk, Artem Huk, Halyna Maistrenko

Abstract. Relevance. For many decades, people have dreamt of creating machines that could match human intelligence, capable of thinking and acting like humans. One of the most fascinating ideas was to enable computers to "see" and interpret the surrounding world in a way that is comprehensible to them. Thanks to progress in the field of artificial intelligence and innovations in deep learning and neural networks, this area has made a significant leap in recent years, surpassing humans in some tasks related to object detection and labeling. **The purpose of this work** is to review and comparatively analyze existing image segmentation tools, particularly those using data clustering methods. **The object of research** is image segmentation algorithms, which are key in the field of computer vision for object recognition and analysis. **The subject of research** is the specific use of clustering methods in these algorithms, their effectiveness, and accuracy. **Results.** This paper conducts a review and comparative analysis of the main methods of data clustering. It highlights 3 of the most common and promising algorithms for improvement. For each of them, the working principle is briefly described, and their main advantages and disadvantages are highlighted. **Conclusion**. Among the methods considered, there is no best universal choice, as the effectiveness of each depends on the specific task. The described algorithms are planned to be used for their further research and modifications.

Keywords: clustering, segmentation, computer vision, image processing, artificial vision.

УДК 004.89 **doi: 10.26906/SUNZ.2024.1.082**

Г. С. Іващенко, Д. О. Тимошенко, О. В. Близнюк, О. М. Кононенко

Харківський національний університет радіоелектроніки, Харків, Україна

МОДЕЛІ ГЛИБОКОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

Анотація. Актуальність. Прогнозування часових рядів є одним із важливих інструментів для різних сфер людської діяльності, оскільки воно дозволяє аналізувати минулі тенденції, розуміти динаміку подій та приймати обґрунтовані рішення на основі попередньо зібраних історичних даних. За останні роки моделі штучних нейронних мереж глибокого навчання показали значний потенціал у сфері прогнозування часових рядів. **Метою даної роботи** є аналіз використання моделей глибокого навчання для короткострокового прогнозування часових рядів різного походження та з можливою наявністю викривлень. **Об'єктом дослідження** є процес прогнозування часових рядів. **Предметом дослідження** є використання моделей глибокого навчання на основі CNN, RNN, TCNN та LSTM архітектур для прогнозування часових рядів. **Результати.** Експериментальні дослідження показали, що прогнози нестаціонарних часових рядів за допомогою штучної нейронної мережі на основі архітектури LSTM виявились найближчими до реальних даних, порівняно з іншими моделями глибокого навчання. **Висновок.** Отримані результати у більшості випадків підтверджують перевагу використання моделей на основі LSTM перед іншими розглянутими моделями глибокого навчання для прогнозування часових рядів.

Ключові слова: прогнозування часових рядів, машинне навчання, штучна нейрона мережа, моделі глибокого навчання, згорткові мережі, довга короткострокова пам'ять.

Вступ

Наразі дані відіграють все більш важливу роль у прийнятті рішень та процесі планування. Поширення аналізу даних, представлених у вигляді часових рядів, відбувається у зв'язку з розвитком сучасних технологій, зокрема, засобів машинного навчання.

Часовий ряд являє собою послідовність спостережень за певним показником у різні моменти часу з однаковим (як правило) інтервалом [1]. Як приклад часових рядів можна привести дані про обсяги продажу продукції за вказаний період або дані про середню температуру повітря на певній території в різні роки. При вирішенні завдання прогнозування часових рядів, в якості вхідних даних можна використовувати не тільки цільовий показник (його минулі значення), а й зовнішні параметри, які можуть прямо або опосередковано впливати на результати прогнозу [1].

Існує багато методів прогнозування часових рядів, але найбільш актуальними є методи на основі підходів машинного навчання [2], таких як генетичні алгоритми, метод висновку за прецедентами, штучні імунні системи та штучні нейронні мережі (ШНМ).

Найбільш поширеним інструментом для прогнозування часових рядів з урахуванням зовнішніх факторів є саме штучні нейронні мережі, через їх можливість враховувати численні параметри для отримання точних прогнозів, а процес налаштування та навчання може бути автоматизовано за допомогою інших засобів обчислювального інтелекту, що забезпечує можливість адаптації моделей прогнозування до нових даних [3]. Перспективним є використання моделей ШНМ глибокого навчання, які дозволяють навчатися для створення прогнозу на викривлених даних, характеризуються можливостями донавчання та урахування зовнішніх факторів, що впливають на величину, яка прогнозується [4]. Існують різні архітектури глибокого навчання, такі як рекурентні нейронні мережі, згорткові нейронні мережі, нейронні мережі з довгостроковою та короткостроковою пам'яттю. Завдяки своїй здатності вивчати складні нелінійні залежності та адаптуватися до різноманітних вхідних даних, ці моделі дозволяють досягти значних результатів у прогнозуванні часових рядів [4].

Однією із проблем вирішення завдання прогнозування часових рядів за допомогою ШНМ є вибір архітектури, з метою чого доцільно розглянути актуальні наукові роботи, у яких широко представлені моделі і методи прогнозування часових рядів на основі штучних нейронних мереж.

У [5] проведено порівняльний аналіз прогнозування часових рядів на основі моделі BiLSTM та статистичними методами, для вивчення розвитку та поширення пандемії COVID-19 в Саудівській Аравії. Відповідно до результатів дослідження, запропонована модель BiLSTM показала значно вищу ефективність відносно використання статистичних методів прогнозування часових рядів. Проте, порівняно з іншими поширеними моделями ШНМ, такими як LSTM (Long Short-Term Memory) і GRU (Gated Recurrent Unit), при прогнозуванні часових рядів на 14 та 60 днів модель BiLSTM показала лише аналогічну точність.

Розглянуте дослідження [6] спрямоване на порівняння ефективності прогнозування за допомогою ШНМ на архітектурі LSTM, MLP (Multilayer Perceptron) та CNN (Convolutional Neural Network). Використаний набір даних щодо спалаху COVID-19 у Єгипті містив записи про підтверджені випадки зараження, смерті та одужання у проміжку від 14 лютого 2020 року до 30 червня 2021 року. 90% даних було використано для навчальної вибірки та 10% для тестової. Виходячи з отриманих результатів, зроблено висновок, що модель на основі архітектури LSTM має найбільшу точність при прогнозуванні часових рядів на тиждень і на місяць уперед. Найгірші результати показала модель CNN, яка під час прогнозування часових рядів з горизонтом, що дорівнює один місяць, почала періодично генерувати аномальні значення.

Виходячи з аналізу сучасних наукових робіт можна зробити висновок, що завдання прогнозування за

допомогою моделей глибокого навчання є актуальним, і потребує детального дослідження.

Метою цієї роботи є аналіз ефективності використання моделей ШНМ глибокого навчання для короткострокового прогнозування часових рядів.

Постановка завдання

Часові ряди являють собою послідовний набір значень, вимірюваних протягом *N* періодів часу,

$$x_t = (x_{t-N}, x_{t-N+1}, x_{t-i}, ..., x_{t-2}, x_{t-1}),$$

де *t* – мітка поточного часу, $0 \le t \le N$ [7].

Вимірювання, зроблені під час процесу, що описується за допомогою часового ряду, розташовані в належному відсортовано-хронологічному порядку.

В дослідженні використання моделей глибокого навчання для прогнозування часових рядів у якості вихідних даних для аналізу обрано дані щодо числа захворілих під час пандемії COVID-19 в Україні (зібрані Всесвітньою Організацією Охорони Здоров'я), дані середньорічної температури повітря за період 2009-2023 років та історичні дані курсів акцій NASDAQ.

Таким чином, для експериментальних досліджень були обрані ряди різного походження, розміру, з відсутніми та наявними зовнішніми факторами та викривленнями.

Дослідження передбачає використання таких моделей ШНМ, як згорткова нейронна мережа (CNN), рекурентна нейронна мережа (RNN), часова згорткова нейронна мережа (TCNN) та довга короткострокова пам'ять (LSTM). Для порівняльного аналізу також використовується багатошаровий персептрон (MLP).

Основна частина

RNN, CNN, TCNN та LSTM – це моделі штучних нейронних мереж, що відносяться до сімейства моделей глибокого навчання. Кожна з моделей може мати три типу шарів: вхідний, прихований і вихідний шар, з'єднані ациклічними зв'язками. Моделі також можуть мати більше одного прихованого шару [8].

Використана архітектура ШНМ на основі RNN для прогнозування часових рядів передбачає декілька фіксованих функціональних блоків активації, по одному на кожний часовий крок [9]. Кожен модуль має внутрішній стан, який визначає попередні знання, які мережа містить на даному етапі. Внутрішній стан оновлюється на кожному лкроці, щоб відобразити зміну знань мережи про минуле на поточному етапі [9].

Прихований стан оновлюється з використанням наступного рекурентного відношення:

$$h(t) = f(h_{t-1}, x_t), \quad (1)$$

де h_t – поточний стан, h_{t-1} – попередній стан, x_t – стан входу. У якості функції активації використовується гіперболічний тангенс зваженої суми:

$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t), \quad (2)$$

де W_{hh} – вага рекурентного нейрона, W_{xh}– вага вхідного нейрона.

Використовувана для прогнозування часових рядів CNN складається з вхідного шару, певної кількості прихованих шарів та вихідного шару [10].

Приховані шари зазвичай складаються зі згорткових шарів, шарів агрегування, нормалізуючих та повнозв'язних шарів. Ці шари пов'язані між собою шарами з визначеними активаційними функціями. Головним елементом згорткової нейронної мережі для прогнозування часових рядів є згорткові шари, де до даних з попереднього шару застосовується операція згортки [10].

Ієрархічна архітектура для прогнозування часових рядів TCNN складається з кількох згорткових прихованих шарів. TCNN використовує три основні техніки: причинні згортки, розширені згортки та обробку залишкових зв'язків.

При використанні техніки причинної згортки вихідні дані в момент часу t згортаються лише з елементами з часу t або більш ранніх часових кроків з попереднього рівня [11]. Нульове доповнення використовується в прихованих шарах, щоб забезпечити, що вони мають ту саму розмірність, що й вхідний шар, щоб спростити згортки.

Методика розширеної згортки дозволяє використати довгу пам'ять, що неможливо лише за допомогою причинних згорток. Розширений згортковий оператор F на елементі послідовності s визначається як:

$$F(s) = \sum_{i=0}^{k-1} f(i) * x_{s-d+1}, \quad (3)$$

де x – це послідовний вхід, k – розмір фільтра, а d – коефіцієнт розширення.

Третя техніка TCNN реалізована за допомогою залишкових блоків [12], що допомагають подолати проблему зникнення градієнта в мережах з багатьма шарами.

Використовувана для прогнозування часових рядів ШНМ на основі архітектури LSTM складається з блоку входу, блоку виходу, вхідного шлюзу та вихідного шлюзу. Комірки пам'яті або комірки LSTM можна розглядати як шар нейронів у традиційній нейронній мережі прямого зв'язку, де кожен нейрон має прихований шар і поточний стан. LSTM має стан комірки, представлений C(t-1) і C(t) для попередньої та поточної міток часу відповідно [13].

Прихований стан реалізує короткочасну пам'ять, а стан комірки використовується у якості довгострокової пам'яті. При використанні комірки штучної нейронної мережі на основі LSTM спочатку визначається збереження інформації з попереднього часового кроку, на підставі:

$$f_t = \sigma(x_t * U_f + H_{t-1} * W_f), \quad (4)$$

де x_t – введення для поточної позначки часу, U_f – вага, пов'язана з введенням, H_{t-1} – прихований стан попередньої позначки часу, W_f – це вагова матриця, пов'язана з прихованим станом [14].

Вхідний шлюз використовується для кількісної оцінки важливості наданої інформації, за допомогою наступного рівняння вхідного вентиля:

$$i_t = \sigma(x_t * U_i + H_{t-1} * W_i), \quad (5)$$

де x_t – введення з поточною міткою часу t, U_i – вагова

матриця введення, H_{t-1} – прихований стан у попередній мітці часу, W_i – вагова матриця введення, пов'язаного з прихованим станом.

Вихідний шлюз ШНМ поєднує поточний вхід, вихід блоку та значення клітинки пам'яті на останній ітерації:

$$O_t = \sigma\left(w_o x_t + r_o y_{t-1} + p_o * c_{t-1} + b_o\right), \qquad (6)$$

де w_o, r_o та p_o – вагові коефіцієнти, b_o – вектор зміщення. Тепер, щоб обчислити поточний прихований стан, використовується O_t і гіперболічних тангенс оновленого стану комірки. Таким чином, прихований стан є функцією довгострокової пам'яті та поточного виведення.

Одним із важливих аспектів удосконалення моделей глибокого навчання є оптимізація гіперпараметрів, яка стосується процесу вибору найкращого набору регульованих параметрів, які контролюють процес навчання моделі та можуть суттєво впливати на її продуктивність [15]. Одними з найпоширеніших методів оптимізації за гіперпараметрами є методи випадкового пошуку, байєсовської оптимізації та генетичні алгоритми.

Результати порівняльного аналізу

Для оцінки прогнозу можуть бути використані численні метрики, такі як середньоквадратична помилка (MSE), середня абсолютна помилка (MAE), середньоквадратичний квадрат помилки (RMSE), середня квадратична помилка прогнозування (MSPE) та інші. Ці метрики мають свої особливості, які важливо враховувати у контексті задачі, оскільки у дослідженні використовується часові ряди різного походження.

MAE вимірює середнє значення абсолютних відхилень, що призводить до меншої чутливості до великих значень помилок у окремих прогнозах.

MAPE визначає відсоткову різницю між фактичними та прогнозованими значеннями, що робить цю метрику корисною для наочної оцінки точності.

MSE вимірює середнє значення квадратів відхилень між фактичними та передбаченими значеннями, і ця метрика доцільна при виявленні аномальних викидів у даних.

Вибір метрики забезпечує розуміння особливостей помилок і недоліків роботи досліджуваної моделі прогнозування часових рядів та її ефективності для конкретного випадку використання.

Для порівняльного аналізу моделей прогнозування на основі архітектур RNN, CNN, LSTM та TCN на першому етапі експериментальних досліджень обраний набір даних, що містить інформацію про розвиток пандемії COVID-19 в Україні. Дані представлені у вигляді більш ніж 135 тисяч записів, кожна з яких складається з 64 полів.

Часові ряди для досліджень можуть містити неповні, відсутні або повторювані дані. Для використання викривлених даних для навчання моделей прогнозування застосовуються підходи щодо попередньої обробки, які полягають у процесі перетворення необроблених даних у значущі за допомогою різних методів.

Першим етапом підготовки вихідних даних часових рядів є усунення пропусків (відсутніх елементів часового ряду).

При навчанні моделей ШНМ наявність пропусків може призводити до неможливості навчання моделі або зниження її точності. При побудові навчальної вибірки за даними стовпця total_tests, який був обраний у якості цільового, можна помітити, що дані не мають значних викидів, а кількість та розташування пропусків дозволяє застосувати поліноміальну інтерполяцію для їх заповнення. Цей метод полягає у знаходженні та використанні полінома найменшого можливого ступеня, який повинен проходити через точки набору даних.

Після першого етапу необхідно застосувати згладжування, для усунення помітних позитивних і негативних викидів. Існує багато методів згладжування, але найбільш поширеними є метод ковзних середніх та експонціальне згладжування. На відміну від методу ковзних середніх, в якому всі дані мають однакову вагу, в методі експоненціального згладжування найбільший коефіцієнт застосовується до останнього спостереження.

Для згладжування часового ряду, в якому є викиди даних, доцільним є метод ковзного середнього, який і був використаний при подальших експериментальних дослідженнях.

Для визначення впливу попередньої обробки даних на якість прогнозування часових рядів, використовувалися нейронні мережі в базовій конфігурації при горизонті прогнозування 20 днів і розмірі історії 60 днів.

Аналіз виконується за допомогою метрики MSE, що розраховується згідно:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \qquad (7)$$

де n – кількість прикладів в наборі даних, y_i – фактичне значення для і-го прикладу, $\hat{y}_i$ – прогнозоване значення для і-го прикладу.

В таблиці 1 представлені результати тестування впливу попередньої обробки даних на результати прогнозування часових рядів, у якості метрики у цьому досліджені обрана MSE.

Таблиця 1 – **Вплив використання попередньої обробки даних на результати прогнозування**

Варіант обробки	Значення Mean Squared Error для моделей прогнозування на основі таких архітектур ШНМ:			
	LSTM	RNN	TCN	CNN
Без попередньої обробки	0,6205	0,5052	0,2805	0,5623
Згладжування new_cases	0,4623	0,4052	0,2001	0,4281
Згладжування new_tests	0,5648	0,4896	0,2249	0,4913
Згладжування new_cases та new_tests	0,4227	0,4006	0,1922	0,4089

Найбільш чутливою до відсутності згладжування виявилася штучна нейронна мережа на основі LSTM. Кращі результати, ніж LSTM, показала ШНМ на основі CNN, але її результати гірші за ШНМ на основі RNN та TCN. Найбільш стійкою до відсутності згладжування є ШНМ на основі TCNN.

У ході дослідження ефективності прогнозування даних щодо поширення COVID-19 було створено 60 варіантів нейронних мереж на основі RNN, CNN, LSTM та TCN архітектур із застосуванням наступних наборів гіперпараметрів:

- для ШНМ на основі LSTM: Units – 16; Activation – tanh; Dropout – 0.4; PredictionHorizon – 20;
- для ШНМ на основі RNN: Units – 64; Activation – tanh; Dropout – 0.4, PredictionHorizon – 20;
- для ШНМ на основі TCN: Filters – 64, Activation – relu, Dropout – 0.3, KernelSize – 3, Dilations – 1, 2, 4, 8, 16, 32; PredictionHorizon – 20;
- для ШНМ на основі CNN: Filters – 64, Activation – relu; Dropout – 0,3; KernelSize – 2; Stride – 1; PredictionHorizon – 20;
- глобальними гіперпараметрами були визначені Epoch – 25 та Batch – 10.

Для візуального аналізу якості прогнозування числових рядів обрано конфігурації нейронних мереж, у яких за результатами експериментів було найменше значення MSE для прогнозування на 20 днів уперед при врахуванні згладжування даних у стовпцях new_tests та new_cases:

- LSTM (Units = 8; Dropout = 0.4, Activation = tanh). Досягнуте значення MSE дорівнює 0,3176;
- RNN (Units = 128; Dropout = 0.4, Activation = tanh). Досягнуте значення MSE дорівнює 0,1103;
- TCNN (Filters = 64; KernelSize = 3; Dilations = 1,4,16,64; Dropout = 0.3). Досягнуте значення MSE дорівнює 0,1090;
- CNN (Filters = 128; KernelSize = 2; KernelStride = 1). Досягнуте значення MSE дорівнює 0,3702.

В якості тестових даних з перевірочного набору були обрані 2 проміжки зі зростаючим і спадним трендом.

Мережі на основі архітектур RNN і TCN врахували можливе зростання захворюваності також як і LSTM, проте прогнозована кількість нових випадків захворювання виявилася вищою за реальні дані.

Необхідність підбору значень гіперпараметрів є однією з проблем при використанні моделей на основі ШНМ глибокого навчання, тому були проведені дослідження впливу гіперпараметрів Units і Dropout на середньоквадратичну помилку прогнозування та час, необхідний для навчання ШНМ, результати наведені у табл. 2.

Проаналізувавши результати тестування впливу гіперпараметрів, можна зробити висновок, що зі збільшенням розмірності вихідного простору (гіперпараметр units) у шарах LSTM-мережи починається поступове збільшення значення MSE і часу навчання. Зміна значення Dropout до 0,8 дозволила уповільнити збільшення MSE та процес перенавчання.

При додаванні нового шару значення MSE збільшується на 0,1, а при зменшенні гіперпараметру units у першому шарі до 8, значення MSE зменшилося до 0,3176.

Подальше зменшення гіперпараметру units у першому шарі призвело до збільшення MSE та втрати точності прогнозування.

Таблиця 2 – **Результати тестування впливу гіперпараметрів на ШНМ на основі LSTM**

Units	Dropout	MSE	Час навчання
2	0,4	0,6467	18
4	0,4	0,4113	18
8	0,4	0,3176	18
16	0,4	0,405	18
32	0,4	0,4397	18
64	0,4	0,5131	36
128	0,4	0,5023	35
64	0,8	0,4037	23

У ході експериментів для аналізу використання архітектури LSTM були також використані історичні дані курсів акцій NASDAQ.

В процесі розрахунків використовувалися техніки масштабування для формування вхідних та вихідних даних для моделі. Метрики помилок прогнозування для отриманих результатів прогнозування на різних кроках у межах горизонту прогнозування наведені у табл. 3.

Таблиця 3 – **Реальні та прогнозовані значення, значення абсолютної та абсолютної у відсотках помилки**

Фактичне	Прогнозоване	AE	APE, %
5220,5	5250,1	29,6	0,56
5039,25	5217,92	178,67	3,54
5086,5	5184,37	97,87	1,92
4936,75	5149,92	213,17	4,31
5224,5	5115	109,5	2,09
5282,75	5080,55	202,2	3,82
5380,5	5047,56	332,94	6,18
5572,5	5017,3	555,2	9,96
5626,5	4991	635,5	11,29
5711,5	4969,81	741,69	12,98
5717,75	4954,63	763,12	13,34
5568,5	4945,88	622,62	11,18
5499	4943,09	555,91	10,10

Наведені розрахунки помилок використані для визначення середніх метрик MAE та MAPE:

$$MAE = \frac{\sum_{i=1}^{k} |y_i - x_i|}{k}, \quad (8)$$

$$MAPE = \frac{100\%}{k} \sum_{i=1}^{k} \left| \frac{y_i - x_i}{y_i} \right|, \quad (9)$$

де k – кількість пар фактичних (y) та прогнозованих (x) значень.

Візуалізація отриманих результатів порівняння фактичних та прогнозованих значень курсів акцій NASDAQ представлена на рис. 1.

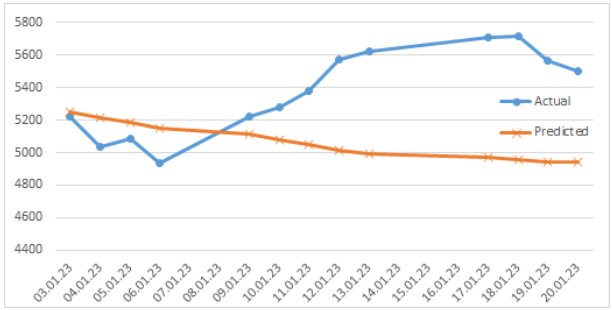


Рис. 1. Результати прогнозування значень курсів акцій NASDAQ за допомогою ШНМ на основі LSTM

Отримані середні значення метрик (MAE – 387,53, MAPE – 7,02) свідчать про успішність застосування нейронних мереж на основі LSTM, принаймні, для короткострокового прогнозу курсів акцій NASDAQ. У ході експериментального дослідження також були використані дані середньорічної температури повітря, прогнозування цих часових рядів є важливою задачею в галузі кліматології та метеорології. Для порівняння були обрані архітектури LSTM та MLP, вхідні дані для аналізу обмежені періодом 2009-2023 років, результати наведені у табл. 4.

Таблиця 4 – **Реальні та прогнозовані значення середньорічної температури (градуси за Фаренгейтом) для ШНМ на основі архітектур MLP та LSTM**

Рік	Фактичне значення	Прогноз (MLP)	Прогноз (LSTM)
2009	54,4	53,651	53,503
2010	53,4	53,738	53,558
2011	51,1	53,607	53,495
2012	53,5	53,177	53,234
2013	52,3	53,517	53,164
2014	53,2	52,995	53,016
2015	54,8	53,497	52,930
2016	54,4	52,998	53,024
2017	54,7	53,159	53,038
2018	53,7	53,567	53,158
2019	53,5	53,102	53,184
2020	53,1	53,282	53,153
2021	54,5	53,901	53,135
2022	54,0	53,478	53,358
2023	53,9	53,934	53,395

Аналізуючи значення метрик ME, MAE, MPE, MAPE, SMAPE, SSE, MSE та RMSE, що наведені у табл. 5, можна визначити, що модель MLP виявилася ефективнішою у короткостроковому прогнозуванні середньорічної температури повітря порівняно з використання моделі глибокого навчання.

Таблиця 5 – **Результати прогнозування часових рядів середньорічної температури**

Метрика помилки	ШНМ на основі MLP	ШНМ на основі LSTM
ME	0,19	0,4
MAE	0,76	0,87
MPE	0,32%	0,73%
MAPE	1,43%	1,63%
SMAPE	1,42%	1,63%
SSE	15,46	18,49
MSE	1,03	1,23
RMSE	1,01	1,11

За результатами дослідження, модель MLP продемонструвала низьку MAE (0,76) та MAPE (1,43%), вказуючи на прийнятну точність прогнозування середньорічної температури повітря. Графічне відображення порівняння прогнозів за допомогою MLP та LSTM наведено на рис. 2.

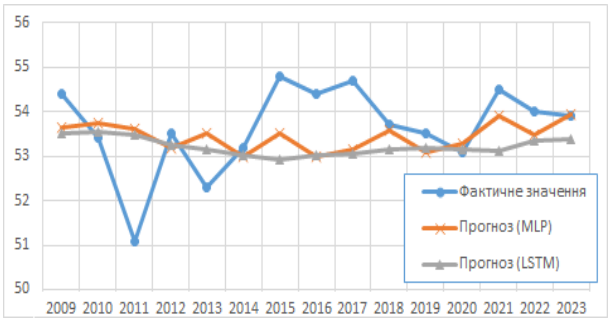


Рис. 2. Результати прогнозування середньорічної температури за допомогою ШНМ на основі LSTM та MLP

Результати експериментальних досліджень вказують на те, що модель на основі MLP виявляється ефективнішою при короткостроковому прогнозуванні необроблених даних середньорічної температури повітря. Це підтверджується низькими значеннями MAE, MAPE, ME та MPE, а також меншими значеннями SMAPE порівняно з LSTM.

З огляду на отримані результати прогнозування даних щодо поширення COVID-19 та курсів акцій NASDAQ можна дійти висновку, що найкращі результати показує мережа на основі архітектури LSTM, але вимагає попередньої обробки вихідних даних. Результати мережі на основі TCNN мають меншу точність в порівняні з результатами LSTM. Мережі RNN та CNN застосовувати недоцільно, бо результати, отримані за допомогою цих мереж, значно відрізняються від реальних даних.

Висновки

В роботі проведений порівняльний аналіз використання моделей штучних нейронних мереж глибокого навчання (LSTM, TCN, CNN, та RNN) для прогнозування часових рядів різного походження: дані щодо поширення COVID-19, історичні дані курсів акцій NASDAQ та часовий ряд середньорічної температури повітря.

У випадку з прогнозуванням числа захворілих під час пандемії COVID-19 найкращий результат показала LSTM мережа, яка має високу ефективність і при прогнозуванні даних NASDAQ. Але запропонована модель прогнозування на основі архітектури LSTM є чутливою до якості вихідних даних, також недостатнє налаштування гіперпараметрів може призводити до значного погіршення точності, що було підтверджено низькими результатами прогнозування середньорічної температури повітря.

Дослідження підтвердило, що моделі глибокого навчання, такі як LSTM, TCNN, CNN, та RNN, можуть ефективно вирішувати завдання прогнозування часових рядів, проте важливо враховувати, що успішність залежить від правильного вибору моделі, оптимальних значень гіперпараметрів, належного навчання та наявності відповідної кількості даних.

Отримані результати можуть служити підґрунтям для розвитку ефективних моделей прогнозування та підтримки прийняття рішень.

Список літератури

1. Chatfield C. (2000), “Time-Series Forecasting”, Chapman and Hall, P. 280.
2. Lin Y., Korsinska Y., Rana M. (2021), “Temporal Convolutional Attention Neural Networks for Time Series Forecasting”, International Joint Conference on Neural Networks, pp. 236-238.
3. Pedro H., Coimbra C. (2012), “Assessment of forecasting techniques for solar power production with no exogenous inputs”, Solar Energy, № 86(7), pp. 2017-2028.
4. Amin Salih Mohammed, Ivashchenko H., Filimonchuk T., Ivanisenko I., Barkovska O. (2020) “Green Hybrid Models Based on Clonal Selection and Case-based Reasoning for Short-term Time Series Forecasting”, Journal of Green Engineering, vol. 10, issue 5, pp. 2139-2154.
5. Shahin A. (2021), “Deep Learning BiLSTM Encoding-Decoding Model for COVID-19 Pandemic Spread Forecasting”, Fractal and Fractional, №4. pp. 175-176.
6. Marzouka M. (2021), “Deep learning model for forecasting COVID-19 outbreak in Egypt”, Process Safety and Environmental Protection, № 153, pp. 363-375.
7. Rangapuram S., Seeger M., Gasthaus D., Stella L., Wang Y, Januschowski T. (2018), “Deep state space models for time series forecasting”, Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), P. 10.
8. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser Ł., Polosukhin I. (2017), “Attention is all you need”, Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), P. 11.
9. Derbentsev V., Matviychuk A., Datsenko N., Bezkorovainyi V., Azaryan A. (2020), “Machine learning approaches for financial time series forecasting”, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy, № 2713, Odessa, Ukraine, pp. 434-450.
10. Hemmati A., Abdoos M., Akbar A. (2015), “Short term load forecasting using a hybrid intelligent method”, Knowledge-Based System, № 76, pp. 139-147.
11. Voyant C., Nivet M., Paoli C., Muselli M., Notton G. (2014), “Meteorological time series forecasting based on MLP modelling using heterogeneous transfer functions”, Journal of Physics: Conference Series, № 574, Madrid, Spain, pp. 28-31
12. Khan S., Rahmani H., Ali Shakh S. A., Bennamoun M (2018), “Guide to Convolutional Neural Networks for Computer Vision”, Morgan & Claypool Publishers, P. 207.
13. Goodfellow I., Bengio Y., Courville A (2016), “Deep Learning (Adaptive Computation and Machine Learning series)”, The MIT Press, P. 775.
14. Greff, K., Srivastava, R., Koutnik, J., Steunebrink, B., Schmidhuber, J. (2017), “LSTM: A Search Space Odyssey”, IEEE Transactions on Neural Networks and Learning Systems, № 28, pp. 2222-2232.
15. Breiman L. (2001), “Random forests”, Machine learning, № 45(1), pp. 5-32.

Received (Надійшла) 11.12.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Deap learning models for time series forecasting

Heorhii Ivashchenko, Daria Tymoshenko, Oleksandr Blyzniuk, Oleksandr Kononenko

Abstract. Topicality. Time series forecasting is one of the important tools for various spheres of human activity, as it allows analyzing past trends, understanding the dynamics of events, and making substantiated decisions based on previously collected historical data. In recent years, deep learning artificial neural network models have demonstrated significant potential in the field of time series forecasting. **The goal of this work** is to analyze the use of deep learning models for short-term forecasting of time series of various origins and with possible presence of distortions. **The object of research** is the process of time series forecasting. **The subject of research** is the use of models based on CNN, RNN, TCNN and LSTM architectures for time series forecasting. **Results.** Experimental research has shown that forecasts of non-stationary time series using an artificial neural network based on LSTM architecture are closer to real data, compared to other deep learning models. **Conclusions.** The obtained results in most cases confirm the advantage of using models based on LSTM over other considered deep learning models for time series forecasting.

Keywords: time series forecasting, machine learning, artificial neural network, deep learning models, convolutional networks, long short-term memory.

УДК 519.7:004.8 **doi: 10.26906/SUNZ.2024.1.088**

І. В. Ільїна[1], В. В. Токарєв[1], А. В. Яковлєв[1], І. І. Шевченко[2]

[1] Харківський національний університет радіоелектроніки, Харків, Україна
[2] Науково-дослідний, проектно-конструкторський та технологічний інститут мікрографії, Харків

ВИКОРИСТАННЯ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОРГАНІЗАЦІЇ ГУМАНІТАРНОЇ ЛОГИСТИКИ

Анотація. У статті досліджується важливість та переваги використання систем підтримки прийняття рішень у волонтерській логістиці, особливо в умовах гуманітарних криз та надзвичайних ситуацій. Сучасний світ стикається з необхідністю швидкого та ефективного реагування на невід'ємні складнощі, що виникають у сфері гуманітарної допомоги. Роль волонтерських організацій у цьому контексті стає визначальною, а організація логістичних процесів вимагає високого рівня координації та точності в прийнятті рішень. Стаття розглядає важливість впровадження сучасних інструментів та технологій у волонтерську логістику і фокусується на використанні систем підтримки прийняття рішень. Детально розглядається алгоритм доставки допомоги волонтерською організацією, починаючи від визначення потреб споживачів до системи моніторингу та збору зворотного зв'язку. Акцент робиться на інноваційних підходах та технологіях, які можуть оптимізувати логістичні процеси волонтерських організацій. Стаття визначає ключові шляхи оптимізації використання систем підтримки прийняття рішень для досягнення найкращих результатів та ефективної гуманітарної допомоги. Наразі основна увага спрямована на створення прототипу системи, який відповідає визначеним критеріям.

Ключові слова: системи підтримки прийняття рішень, DSS, волонтерські організації, логістика.

Вступ

Постановка проблеми. В сучасному світі, де гуманітарні кризи та надзвичайні ситуації стають невід'ємною частиною нашого життя, роль волонтерських організацій у сфері логістики стає дедалі важливішою. Забезпечення швидкої та ефективної допомоги у потрібний момент вимагає високого рівня організації та точності при прийнятті рішень. У цьому контексті використання систем підтримки прийняття рішень набуває ключового значення для координації дій волонтерів у логістичних процесах.

Ця стаття присвячена розгляду важливості впровадження сучасних інструментів та технологій у сферу волонтерської логістики. Аналізуючи вплив систем підтримки прийняття рішень на організацію, управління та забезпечення оптимізації логістичними процесами, розкриваються переваги їх використання у забезпеченні ефективності та оперативності допомоги під час надзвичайних ситуацій [1, 2]. Треба відмітити роль інноваційних підходів у забезпеченні логістичної підтримки волонтерських ініціатив та визначити шляхи оптимізації використання цих систем для досягнення найкращих результатів.

Доставка гуманітарної допомоги волонтерською організацією проводиться у такій послідовності:

1. Волонтерська організація розпочинає процес, визначаючи потреби кінцевих споживачів та виділяючи конкретні товари або послуги, які потрібно доставити;

2. Залучення та реєстрація волонтерів, які виразили бажання допомагати у транспортуванні замовлень;

3. Групування та категоризація замовлень для оптимального використання ресурсів волонтерів. Можливо, виділення пріоритетних або термінових доставок;

4. Використання алгоритмів маршрутизації для оптимізації маршрутів волонтерів та забезпечення ефективності транспортування. Це може включати врахування географічного розташування споживачів та трафіку.

5. Забезпечення ефективного зв'язку між волонтерами, організацією та споживачами. Надання волонтерам необхідної інформації про замовлення та надання підтримки в разі виникнення питань або проблем.

6. Встановлення системи моніторингу для відстеження статусу замовлень та роботи волонтерів.

На сьогоднішній день основні сили спрямовані на створення прототипу системи прийняття рішення, який буде відповідати заданим критеріям.

Мета статті – розгляд алгоритму доставки допомоги волонтерською організацією: від кроку визначення потреб кінцевих споживачів до збору зворотного зв'язку для постійного удосконалення системи. Висвітлення ролі системи підтримки прийняття рішень у кожному кроці алгоритму допоможе виділити їх важливість у забезпеченні ефективності та оперативності допомоги.

Виклад основного матеріалу

Система підтримки прийняття рішень (СППР, англ. Decision Support System, DSS) – це інформаційна система, яка підтримує діяльність з прийняття бізнес- або організаційних рішень. СППР обслуговують рівень управління, операцій і планування організації (зазвичай керівництво середньої та вищої ланки) і допомагають людям приймати рішення щодо проблем, які можуть швидко змінюватися та нелегко визначити заздалегідь, тобто неструктурованих і напівструктурованих проблем прийняття рішень. Системи підтримки прийняття рішень можуть бути або повністю комп'ютеризованими, або керованими людиною, або поєднувати обидва варіанти [3]. Хоча науковці сприймають DSS як інструмент для підтримки процесів прийняття рішень, користувачі бачать DSS як інструмент для полегшення організаційних процесів [4].

Алгоритм доставки волонтерської допомоги, який представлений умовною схемою (рис. 1), виглядає таким чином:

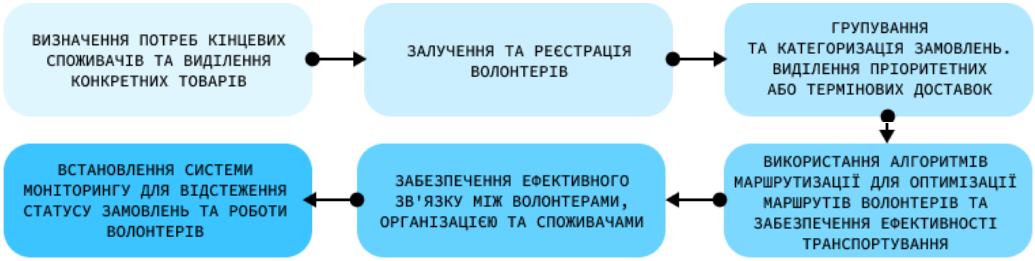


Рис. 1. Схема алгоритму доставки волонтерської допомоги

Крок 1. Визначення потреб кінцевих споживачів. Група дослідників Jack Canfield, Mark Victor Hansen, Les Hewitt [5] запропонували 4 правила тайм-менеджменту, які можуть бути використані для організації волонтерських запитів: Do – це завдання, які потрібно зробити зараз; Drop – ці речі не впливають на роботу волонтера чи організації; Defer – це завдання, які слід виконати зрештою; Делегувати – це елементи, які інші волонтерські організації можуть допомогти виконати. Це критично важливий навик для хорошого менеджменту.

Крок 2. Залучення та реєстрація волонтерів. Реєстрація волонтерів буде виконуватися за допомогою web-додатку і складатися з наступних кроків: аутентифікація за допомогою логін/пароль або постачальника аутентифікації (Google, Facebook, та ін.); підтвердження особи. Для цього волонтер повинен буде надіслати документи такі як паспорт, фотографію з паспортом, ідентифікаційний код, посвідчення волонтера, якщо є. Також можливе проведення онлайн співбесіди, це необхідно для уникнення шахрайства. Після підтвердження особи адміністратором особа може бути допущена до виконання волонтерських замовлень. Залучення до виконання замовлення може бути організоване наступним чином: визначення відповідального за доставку замовлення; залучення необхідної кількості волонтерів для виконання завдання. Цей крок також може бути зроблений за допомогою web-додатка.

Крок 3. Групування та категоризація замовлень. У більшості проблем, варіанти рішення необхідно оцінювати з різних точки зору, враховуючи фізичні, економічні, технічні та інші параметри, що впливають на складність вирішення задачі. Формула пріоритезації волонтерських замовлень може бути складною і визначатися відповідно до конкретних цілей та пріоритетів волонтерської організації. Проте, основні фактори, які можна враховувати при розробці такої формули, включають: терміни; специфікації та об'єм замовлення; відстань; ефективність використання ресурсів; важливість замовлення; доступність волонтерів. Враховуючи ці фактори, було запропоновано формулу, яка призначатиме числові значення кожному замовленню (для кожного фактору визначена певна шкала, наприклад від 1 до 10) та допомагатиме визначити його пріоритет K_i у черзі. Тоді формула (1) набуває такого вигляду:

$$K_i = \frac{0.4 \times I_i + 0.3 \times \mathrm{T}_i + 0.2 \times A_i + 0.1 \times E_i}{S_i + D_i}, \quad (1)$$

де I – Важливість, T – Терміни, E – Ефективність, A – Доступність, S - Специфікації, D – Відстань.

Крок 4. Використання алгоритмів маршрутизації для оптимізації маршрутів волонтерів. Для оптимізації маршрутів у відповідності "Природні обчислення" використовуються генетичні алгоритми, які враховують принципи природних механізмів прийняття рішень [6, 7]. Генетичні алгоритми дозволяють швидко генерувати рішення для задачі маршрутизації, уникаючи повного перебору та значно зменшуючи часові витрати. Традиційні методи оптимізації і пошуку рішення можуть вимагати повторного виконання всіх обчислень при навіть невеликій зміні параметрів середовища передавання даних. Генетичні алгоритми ґрунтуються на принципах природної еволюції та стратегії «виживає найпристосованіший». Найкращі рішення виживають і змінюються до досягнення оптимального маршруту передавання даних. Мурашкові алгоритми є імовірнісною жадібною евристикою, де ймовірності встановлюються на основі інформації про якість попередніх рішень. У модифікованому мурашковому алгоритмі для оптимальної маршрутизації, запропонованому в роботах [8, 9], використовується покращена евристика мурашки за рахунок введення поправки на кут.

Але прийняття рішень командою волонтерів у завдання транспортної логістики воєнного часу, тобто в умовах динамічних змін зовнішнього середовищі, має такі особливості: більшість рішень приймаються в ситуаціях, які раніше не зустрічалися; вибір варіантів рішення відбувається в умови неповної та невизначеної інформації про поточну ситуацію; рішення, як правило, найбільш відповідальні проводиться в жорстких часових обмеженнях і постійної зміни інформації. Тож для кожної конкретної доставки може бути обраний власний алгоритм, або на основі робот вище приведених авторів може бути розроблений універсальний алгоритм.

Крок 5. Забезпечення ефективного зв'язку між волонтерами, організацією та споживачами. Ефективний зв'язок між волонтерами, організацією та споживачами є ключовим елементом успішної роботи волонтерської організації. Це забезпечує координацію дій, підвищує ефективність виконання завдань та сприяє покращенню загального враження від послуг.

Для забезпечення ефективної комунікації можуть бути використані: електронна пошта для розсилки важливих інформаційних повідомлень волонтерам, організаціям та споживачам; месенджери або платформи для миттєвих повідомлень (instant messages) для швидкого обміну короткими повідомленнями між волонтерами та організацією; інформаційний портал або web-сайт, на якому можна

розміщувати важливі оголошення, розклади подій, інструкції та іншу інформацію; групи у соціальних мережах для взаємодії волонтерів, обговорення питань та обміну ідеями; електронні опитування для забезпечення зв'язку та отримання зворотної інформації від волонтерів та споживачів. Забезпечення ефективного зв'язку вимагає комплексного підходу та використання різноманітних інструментів, щоб забезпечити взаємодію між всіма сторонами і покращити загальний результат волонтерської діяльності.

Крок 6. Встановлення системи моніторингу для відстеження статусу замовлень та роботи волонтерів. Для таких цілей може бути імплементована в web-додаток CRM система. CRM – модель взаємодії, яка визначає, що центром всієї філософії бізнесу є клієнт, а основними напрямами діяльності є заходи з підтримки ефективного маркетингу, продажу та обслуговування клієнтів. Підтримка цих бізнес-цілей включає збір, збереження та аналіз інформації про споживачів, постачальників, партнерів, а також про внутрішні процеси компанії [10]. Також до цієї системи можна додати трекінг замовлення щоб клієнт міг відстежувати статус замовлення.

Висновки

У статті обґрунтовано важливість впровадження сучасних інструментів та технологій у сферу волонтерської логістики, висвітлена роль систем підтримки прийняття рішень у забезпеченні ефективності та оперативності допомоги в умовах гуманітарних криз та надзвичайних ситуацій.

Слід зазначити, що використання таких систем допомагає координувати дії волонтерів та оптимізувати логістичні процеси, що є ключовим у вирішенні невідкладних потреб.

Алгоритм доставки волонтерською організацією, який було розглянуто, підкреслив важливість кожного кроку в організації та забезпеченні логістичної підтримки. Використання інноваційних підходів, таких як алгоритми маршрутизації та системи моніторингу, сприяє оптимізації роботи волонтерів та покращує комунікацію волонтерів, організації та споживачів.

Розглянутий алгоритм в подальшому буде використаний для створення прототипу системи «Волонтерська організація», який відповідає визначеним критеріям.

Загальний висновок полягає в тому, що використання систем підтримки прийняття рішень у волонтерській логістиці є необхідним елементом для забезпечення ефективності та успішності допомоги в умовах складних ситуацій. Досягнення найкращих результатів вимагає поєднання інновацій, організаційної точності та взаємодії всіх учасників процесу.

Список літератури

1. Koshevoy N., Ilina I., Tokariev V., Malkova A., Muratov V. Implementation of The Gravity Search Method For Optimization By Cost Expenses Of Plans For Multifactorial Experiments, Radioelectronic and Computer Systems, 2023, №. 1(105), pp. 23-32.
2. Ruban, I., Ilina, I., Mozhaiev, M. Researching priority directions in the area of Data, Control navigation and communication systems, 2020, no. 4(62), pp. 59-63. DOI: 10.26906/SUNZ.2020.4.059.
3. Bandyopadhyay S. Decision Support System: Tools and Techniques (1st ed.). CRC Press, 2023. P. 394. doi.org/10.1201/9781003307655
4. Sprague, Ralph H. A Framework for the Development of Decision Support Systems. MIS Quarterly, vol. 4, no. 4, 1980, pp. 1–26. JSTOR. doi.org/10.2307/248957. Accessed 13 Jan. 2024.
5. Canfield J., Hansen M., Hewitt L. The Power of Focus Tenth Anniversary Edition: How to Hit Your Business, Personal and Financial Targets with Absolute Confidence and Certainty Paperback, Pub: Health Comm. Inc; Anniversary ed, 2012, P. 384.
6. Стеценко, І.В. Моделювання систем: навч. посіб. [Електронний ресурс, текст] / І.В. Стеценко; М-во освіти і науки України, Черкас. держ. технол. ун-т., Черкаси: ЧДТУ, 2010. 399 с.
7. Katoch, S., Chauhan, S.S., Kumar, V. A review on genetic algorithm: past present and future. Multimedia Tools and Applications, 2021, vol. 80, Issue 18, pp. 8091-8126. DOI: 10.1007/s11042-020-10139-6.
8. Skakalina E. Застосування мурашиних алгоритмів в рішенні задачі маршрутизації / E. Skakalina // Системи управління, навігації та зв’язку. Полтава: ПНТУ, 2019. – Т. 6 (58). – С. 75-83. – doi:https://doi.org/10.26906/SUNZ.2019.6.075.
9. Галелюка І.Б. Моделювання бездротових сенсорних мереж. Комп’ютерні засоби, мережі та системи. 2015. № 14. С. 141–150.
10. Петецкі І, Крикавський Є, Гладій У, Черкес Р. Актуальність впровадження CRM-систем на підприємствах. AV [інтернет]. 03, Квітень 2023. доступний у: https://academy-vision.org/index.php/av/article/view/298.

Received (Надійшла) 11.12.2023
Accepted for publication (Прийнята до друку) 07.02.2024

Using a decision support system for organizing humanitarian logistics

I. Ilina, V. Tokariev, A. Yakovliev, I. Shevchenko

Abstract. The article explores the importance and advantages of using decision support systems in volunteer logistics, especially in the context of humanitarian crises and emergencies. The modern world faces the necessity for quick and effective responses to inherent difficulties in the field of humanitarian aid. The role of volunteer organizations becomes crucial in this context, and the organization of logistics processes requires a high level of coordination and precision in decision-making. The article discusses the importance of implementing modern tools and technologies in volunteer logistics and focuses on the use of decision support systems. The delivery algorithm by a volunteer organization is detailed, starting from identifying consumer needs to the monitoring system and feedback collection. The emphasis is on innovative approaches and technologies that can optimize logistics processes for volunteer organizations. The article identifies key ways to optimize the use of decision support systems to achieve the best results and efficient humanitarian aid. Currently, the primary focus is on creating a prototype system that meets specified criteria.

Keywords: decision support system, DSS, volunteer organizations, logistics.

УДК 629.3 **doi: 10.26906/SUNZ.2024.1.091**

С. І. Клівець, О. В. Кулєшов, Т. В. Кулєшова

Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна

МЕТОД ВИЗНАЧЕННЯ ЗМІНИ СТАНІВ СКЛАДНОЇ ГЕТЕРОГЕННОЇ СИСТЕМИ ПРИ ОПЕРАТИВНОМУ УПРАВЛІННІ

Анотація. **Актуальність.** При оперативному управлінні необхідно прогнозувати можливі зміни станів складної системи. Але при суттєвій гетерогенності блоків складної системи стандартні методи розрахунку станів надають прогноз з неприпустимими похибками. **Мета статті** – розробити метод визначення зміни станів складної гетерогенної системи з припустимими похибками при оперативному управлінні. **Результати дослідження**. Розроблена математична модель зміни станів складної гетерогенної системи. Технічний стан блока в будь-який момент часу в моделі характеризується агрегованим параметром, який є ймовірністю безвідмовної роботи блока. На основі розробленої моделі запропонований метод побудови функції перехідних ймовірностей процесу зміни значень агрегованого параметра. **Висновок.** Розроблений метод дозволяє зменшити похибку прогнозу зміни стану складної гетерогенної системи при введенні агрегованого параметра.

Ключові слова: складна система, гетерогенність, стан системи, оперативне управління.

Вступ

Однією з основних проблем при оперативному управлінні є питання прийняття рішень в стохастичних системах. Такі проблеми, наприклад, виникають в різноманітного роду системах управління складними об'єктами. Є два напрямки розвитку методів прийняття оптимальних рішень в стохастичних системах [1]. Одне з них тісно примикає до нелінійного програмування. Це стохастичне програмування, в якому розглядаються задачі умовного екстремуму при наявності додаткових обмежень, що враховують стохастичну природу систем [2]. Тут начальне рішення інтерпретується як деякий план одноразового застосування. Але цей напрямок мало пристосований для застосування як апарат для прийняття рішень в задачах оперативного управління, хоча така можливість не виключається [3, 4]. Другий напрямок орієнтований на динамічні, керуємі і стохастичні системи. Методи прийняття рішень в них ґрунтуються на управлінні станом, що може змінюватися під впливом керувань деяким випадковим чином. Очевидно, що це направлення природним чином пристосоване для прийняття рішень в задачах оперативного управління [5, 6]. Але при такому підході при суттєвій гетерогенності блоків складної системи стандартні методи розрахунку станів надають прогноз з неприпустимими похибками [7–10].

Мета роботи – розроблення методу визначення зміни станів складної гетерогенної системи з припустимими похибками при оперативному управлінні.

1. Математична модель зміни станів складної гетерогенної системи

Для завдання моделі марківського процесу рішення треба визначити:

– множину станів процесу, що досліджується;

– множину керуючих впливів, що застосовуються в гетерогенній системі, яка знаходиться в одному з можливих станів;

– ймовірності переходів системи з одних станів складної гетерогенної системи в інші при застосуванні різноманітних керуючих впливів;

– безпосередній виграш при переході складної гетерогенної системи з одного стану в інший.

При математичній формалізації перерахованих компонентів моделі найбільшу складність, як правило, викликає завдання третьої компоненти – перехідних ймовірностей досліджуваного процесу еволюції системи.

Така проблема виникає, наприклад, при визначенні оптимальної стратегії управління технічним станом складних об'єктів з континуальною множиною станів [10].

Нехай є складна гетерогенна система управління складним об'єктом. Момент надходження заявки на використання даного об'єкта є випадковим, тобто заздалегідь невідомим. При використанні складного об'єкта за призначенням у випадковий момент часу t система управління повинна безвідмовно відпрацювати якійсь конкретний час T_p. Якщо система управління складається з K блоків ($k = 1, K$), то при використанні об'єкта кожний з її блоків буде відпрацьовувати час T_p^k.

Технічний стан k-го блока в будь-який момент часу будемо характеризувати агрегованим параметром $z^k(t)$, який є ймовірністю безвідмовної роботи блока протягом часу T_p^k, при умові. що заявка на застосування об'єкта надійде в момент часу t. Тоді стан системи управління буде однозначно (з точки зору надійності) визначатися сукупністю таких агрегованих параметрів блоків:

$$z(t) = (z^1(t), z^2(t), ..., z^k(t), ..., z^K(t)). \quad (1)$$

Таким чином, параметр z приймає своє значення із множини S, яка має потужність, що описується таким декартовим добутком:

$$|S| = \left[0, z_b^1\right] \times \left[0, z_b^2\right] \times ... \times \left[0, z_b^k\right] \times ... \times \left[0, z_b^K\right], \quad (2)$$

де z_b^k – максимально значення агрегованого параметра k-го блока, що володіє конкретною структурною схемою надійності і не має відмов за час m_p^k.

Кожний блок, що входить в склад системи управління, повинен проходити контролювання, при цьому одна частина блоків періодично перевіряється тестовим контролем з глибиною до елемента, що контролюється а інша частина блоків наражається на постійний апаратурний контроль, на працездатність. При цьому в силу неідеальності системи контролю виникають помилки контролю першого і другого роду. Під час роботи системи управління при надходженні заявки також проводиться контроль кожного блока, однак повнота цього виду контролю менше, тому помилки першого і другого роду в цьому випадку мають більше значення.

Якщо в момент t_0 агрегований параметр k-го блока мав значення

$$z^k(t_0) = z_0^k, \tag{3}$$

то в момент $t_0 + t$ він може прийняти або значення

$$z^k(t) = z_t^k = 0, \tag{4}$$

або одне із значень, що належить відрізку

$$\left[z_н^k, z_t^к \right], \tag{5}$$

де $z_н^k$ – нижнє значення агрегованого параметра, при якому k-й блок гарантовано здатний безвідмовно проробити час T_p^k.

Розглянемо різницю

$$\eta^k(t) = z_t^k - z_0^k. \tag{6}$$

В [1] показано підхід до визначення перехідних імовірностей абстрактного полумарківського процесу зміни значень деякого параметра z при відомій умовній функції розподілу такого вигляду:

$$F_k = F_{\eta^k}\left(v \middle| z_0^k, t\right). \tag{7}$$

Використовуючи цей підхід, перехідні ймовірності процесу зміни значень агрегованого параметра можна подати у вигляді $\left(c^k \prec b^k\right)$:

$$q^k\left(t, \left(b^k, c^k\right] \middle| z_0^k\right) =$$

$$= \begin{cases} F_{\eta^k}\left(z_0^k - c^k \middle| z_0^k, t\right) - \\ -F_{\eta^k}\left(z_0^k - b^k \middle| z_0^k, t\right), & z_0^k > b^k > c^k,\ z_0^k \geq z_p^{\ell}; \\ F_{\eta^k}\left(z_0^k - c^k \middle| z_0^k, t\right), & b^k \geq z_0^k > c^k,\ z_0^k \geq z_н^k; \\ 0, & b^k > c^k \geq z_0^k,\ z_0^k \geq z_н^k; \\ 0, & c^k \neq 0,\ z_0^k < z_н^k; \\ 1, & c^k = 0,\ z_0^k < z_н^k. \end{cases} \tag{8}$$

2. Метод визначення зміни агрегованого параметра складної гетерогенної системи

Розглядувана математична модель (1)–(8) призначена для визначення аналітичного виразу умовної функції розподілу F_k при дослідженні конкретного процесу зміни агрегованого параметра заданої складної гетерогенної системи.

Вираз для F_k є функцією 3-х аргументів V, $z_0^к$, t. Метод полягає в багатократному моделюванні процесу експлуатації системи управління таким чином, щоб кожному відомому поєднанню значень аргументів ставало у відповідність значення функції. Таким чином визначається функція таблиця дискретних значень F_k. Далі шляхом використання стандартних програм інтерполяції, наведених, наприклад, в [9], визначається аналітичний вираз цієї функції.

Розглянемо покроковий алгоритм отримання дискретних значень функції F_k.

Крок 1. Визначити значення $z_в^к$, $z_н^k$. Ці значення визначаються одним з відомих методів [6] для заданих структурної схеми надійності та інтенсивності відмов комплектуючих блок елементів.

Крок 2. Покласти $l = 1$.

Крок 3. Випадковим чином задати вхідний l-й стан блока в момент $t_0 = 0$.

Крок 4. Шляхом імітації процесу роботи блока визначити для моменту $\mathbf{t_0}$ значення агрегованого параметра блока, який знаходиться в l-му стані

$$z_0^{(l)} = P_{S_0^{(l)}}(T_P^k).$$

При цьому необхідно врахувати характеристики системи контролю (повноту, помилки контролю).

Крок 5. Покласти $c = 0$.

Крок 6. Вибрати значення $t = t_c$.

Крок 7. Покласти $n = 1$, $r_u = 0$.

Крок 8. Змоделювати спонтанну еволюцію блока на відрізку часу $[t_0, t]$.

Крок 9. Моделюючи процес роботи блока визначити для моменту t значення агрегованого параметра блока

$$z_{kn}^{(l)} = P_{S_{kn}^{(l)}}(T_P^k)$$

з урахуванням характеристик системи контролю.

Крок 10. Обчислити значення

$$\eta_{кп}^{(l)} = z_0^{(l)} - z_{кп}^{(l)}.$$

Крок 11. Покласти $u = 1$.

Крок 12. Обчислити різницю $v_u = z_в - z_н$.

Крок 13. Перевірити умову

$$\eta_{кп}^{(l)} \leq v_u.$$

Якщо "так", то перейти до кроку 14, а в протилежному випадку перейти до кроку 16.

Крок 14. Покласти $r_u = r_u + 1$.

Крок 15. Обчислити значення

$$F_{\eta^{k(l)}}\left(V_m \middle| z_0^{(l)}, t_k\right) = \frac{r_u}{N}.$$

Крок 16. Перевірити умову $u < U$. Якщо "так", то перейти до кроку 17. В протилежному випадку перейти до кроку 19.

Крок 17. Покласти $u = u + 1$.

Крок 18. Обчислити значення $v_u = v_u - \frac{v_u}{2 - u/U}$ та повернутися до кроку 13.

Крок 19. Перевірити умову $n < N$. Якщо "так", то перейти до наступного кроку. В протилежному випадку треба перейти до кроку 21.

Крок 20. Покласти $n = n + 1$ і перейти до кроку 8.

Крок 21. Запам'ятати поточні значення $z_0^{(l)}$, t_k, v_u, $F_{\eta_k^{(l)}}\left(V_u z_0^{(l)}, t_k\right)$.

Крок 22. Перевірити умову $k < K$. Якщо "так", то перейти до наступного кроку. В протилежному випадку треба перейти до кроку 24.

Крок 23. Покласти $k = k + 1$ і перейти до кроку 6.

Крок 24. Перевірити умову $l < L$. Якщо "так", то перейти до наступного кроку. В протилежному випадку треба перейти до кроку 26.

Крок 25. Покласти l = l + 1 і перейти до кроку3.

Крок 26. Виведення вихідних даних і запуск стандартної програми апроксимації функції трьох аргументів. Кінець.

Вибравши достатнє значення U, одержують аналітичні вирази (8) для перехідних імовірностей полумарківського процесу зміни агрегованого параметра блока. Різноманітні сукупності елементів декартового добутку таких перехідних імовірностей дадуть вираз для ймовірностей переходів процесу зміни агрегованого параметра системи управління в цілому.

Висновки

В результаті проведених досліджень розроблена математична модель зміни станів складної гетерогенної системи. Технічний стан блока в будь-який момент часу в моделі характеризується агрегованим параметром, який є ймовірністю безвідмовної роботи блока. Розроблена модель враховує зміни технічного стану апаратури з урахуванням характеристик системи контролю і відновлення. На основі розробленої моделі запропонований метод побудови функції перехідних ймовірностей процесу зміни значень агрегованого параметра.

Розроблений метод дозволяє зменшити похибку прогнозу зміни стану складної гетерогенної системи при введенні агрегованого параметра.

Список літератури

1. Emzir, M.F. Efficient projection filter algorithm for stochastic dynamical systems with correlated noises and state-dependent measurement covariance. Signal Processing, 2024. 218, 109383. Doi: https://doi.org/10.1016/j.sigpro.2024.109383
2. Chetthamrongchai, P., Sayed, B.T., Artemova, E.I., Bashar, B.S., Heri Iswanto, A. Designing a Mathematical Model to Solve the Uncertain Facility Location Problem Using C Stochastic Programming Method. Foundations of Computing and Decision Sciences, 2023. Vol. 48(3). Pp. 345–355. Doi: https://doi.org/10.2478/fcds-2023-0014
3. Pratama, I.N., Dachyar, M., Pratama, N.R. Optimization of Resource Allocation and Task Allocation with Project Management Information Systems in Inf. Techn. Companies. TEM Journal. 12(3). Pp. 1814–1824. https://doi.org/10.18421/TEM123-65
4. Kuchuk G., Kovalenko A., Komari I.E., Svyrydov A., Kharchenko V. Improving big data centers energy efficiency: Traffic based model and method. Studies in Systems, Decision and Control, vol 171. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (Eds.). Springer Nature Switzerland AG, 2019. Pp. 161-183. DOI: http://doi.org/10.1007/978-3-030-00253-4_8
5. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об'єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: https://doi.org/10.20998/2522-9052.2018.1.04
6. Nechausov A., Mamusuĉ I., Kuchuk N. Synthesis of the air pollution level control system on the basis of hyperconvergent infrastructures. *Сучасні інформаційні системи*. 2017. Т. 1, № 2. С. 21 – 26. DOI: https://doi.org/10.20998/2522-9052.2017.2.04
7. Iervolino, R., Manfredi, S. Global stability of multi-agent systems with heterogeneous transmission and perception functions. Automatica. 2024. 162, 111510. DOI: http://doi.org/10.1016/j.automatica.2024.111510
8. Кучук Н. Г., Мерлак В. Ю., Скородєлов В. В. Метод зменшення часу доступу до слабкоструктурованих даних. *Сучасні інформаційні системи*. 2020. Т. 4, № 1. С. 97-102. doi: https://doi.org/10.20998/2522-9052.2020.1.14
9. Kovalenko, A. and Kuchuk, H. (2022), "Methods to Manage Data in Self-healing Systems", Studies in Systems, Decision and Control, Vol. 425, pp. 113–171, doi: https://doi.org/10.1007/978-3-030-96546-4_3
10. Shi, H., Lin, W., Liu, C., Yu, J. A Novel Heterogeneous Parallel System Architecture Based EtherCAT Hard Real-Time Master in High Performance Control System. Electronics. 11(19), 3124. Doi: http://doi.org/10.3390/electronics11193124

Received (Надійшла) 29.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Method of determining the change of state of a complex heterogeneous system under operational management

Sergii Klivets, Alexander Kuleshov, Tetiana Kulieshova

Abstract. Topicality. During operational management, it is necessary to predict possible changes in the state of a complex system. But with significant heterogeneity of the blocks of a complex system, standard methods of calculating states provide a forecast with unacceptable errors. **The purpose of the article** is to propose a method for determining the state change of a complex heterogeneous system with acceptable errors during operational control. **Research results.** A mathematical model of the change of states of a complex heterogeneous system has been developed. The technical state of the block at any moment in time in the model is characterized by an aggregated parameter, which is the probability of the block's failure-free operation. On the basis of the developed model, a method of constructing the function of transitional probabilities of the process of changing the values of the aggregated parameter is proposed. **Conclusion.** The developed method makes it possible to reduce the forecast error of a change in the state of a complex heterogeneous system when an aggregated parameter is introduced.

Keywords: complex system, heterogeneity, system state, operational management.

UDC 629.3 **doi: 10.26906/SUNZ.2024.1.094**

Zakhar Kolesnyk[1], Oleksandr Mezhenskyi[1], Oleksandr Davykoza[1], Heorhii Kuchuk[2]

[1] Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
[2] National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

FOG COMPUTING TECHNOLOGY IN DISTRIBUTED SYSTEMS

Abstract. Topicality. The concept of fog computing is an evolutionary stage in the development of the cloud concept. It occupies a leading position among the general trends in the development of information technology. The emergence of this concept is closely related to the origin and development of the concept of the Internet of Things. **The results.** The subject area was analyzed. It includes an analysis of current trends in the field of organizing distributed computing, an analysis of the use of population algorithms and ontology models for solving optimization problems in distributed systems, an analysis of models, methods and algorithms for solving the problem of transferring the computational load in distributed systems implemented on the basis of fog computing. **Conclusion.** It has been revealed that the concept of fog computing makes it possible to solve most of the problems associated with the load on the communication infrastructure and the latency of information exchange. But they do not resolve issues related to the high dynamism of the foggy environment and the concomitant decrease in the efficiency of the distributed system.

Keywords: computer system, distributed system, fog computing, cloud computing, Internet of Things.

Introduction

Currently, the new paradigm of fog computing, which can be considered as an extension of the cloud concept, has found wide application in many fields. This is due to the fact that today clouds do not meet the high requirements of mobility, low latency and local awareness [1, 2]. A promising solution in this case seems to be a selective shift of computing, communication, control and decision making to the places where data is generated, which corresponds to the basic principles of the concept of fog computing [3, 4].

The concept of fog computing is an evolutionary stage in the development of the cloud concept. It occupies a leading position among the general trends in the development of information technology. The emergence of this concept is closely related to the origin and development of the concept of the "Internet of Things" (IoT) [5, 6]. Scenarios for using the concept of fog computing are very diverse. They are determined by the development of related technologies. This concept has been successfully used in the creation of systems such as smart home, smart transport, e-health, e-government, trade and financial services, industrial production, technological and business process management, and much more [7–12].

The definition of "fog computing" was first introduced by Cisco in 2011.

ІІІщЕ and is currently actively developing due to the following factors:

– the computing power of communication equipment and devices located in the fog layer allows for additional calculations;

– the number of end devices is growing very quickly, and this trend will continue in the foreseeable future.

– it is advisable to relieve the data processing center (including cloud ones) from performing complex computing processes [13, 14].

Fog computing is a tiered model that provides ubiquitous access to a shared pool of scalable computing resources. Fog computing minimizes the network response time of supported applications and also provides end devices with local computing resources and, if necessary, network connectivity to centralized services" [15]. The fog computing architecture can be considered as a "layer" between the cloud and end (user) devices (Fig. 1).

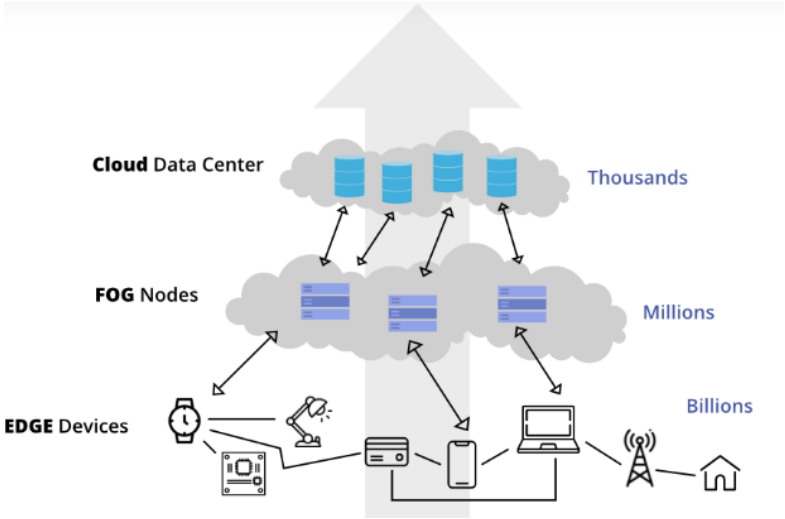


Fig. 1. Fog computing concept

The fog computing environment consists of network components such as routers, switching devices, TV converters, proxy servers, base stations, sensors, etc.

The key difference between the fog concept and the cloud concept is the ability to dynamically transfer the computing load from the cloud to the periphery of the network infrastructure using fog layer devices, as well as partial placement of the load in the fog layer.

This allows you to significantly reduce the load on the communication infrastructure of the network [16].

The purpose of the article is to determine the features of fog computing in distributed systems

Research results

The main problems associated with the organization of the Internet of Things concept and the use of fog computing technology to solve them are shown in Fig. 2.

Major cloud providers: Amazon, Google and Microsoft are developing the concept of fog computing based "serverless architecture"

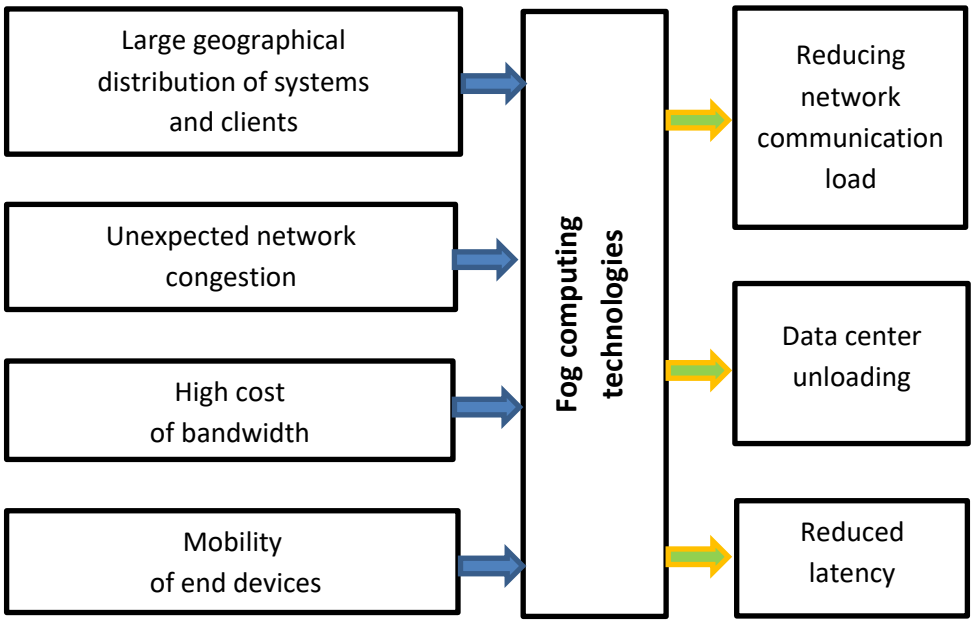


Fig. 2. Prospects for using fog computing technology to solve problems associated with the organization of the Internet of Things

Amazon offers the AWS IoT Greengrass fog computing platform, which effectively extends cloud infrastructure to fog layer devices, allowing data to be processed locally while using the cloud to manage, analyze and store the data. Such a platform allows you to programmatically filter device data and transmit only the necessary information back to the cloud [15]. Microsoft has proposed a solution for serverless computing – Azure Functions, which allows you to run small fragments of program code, or functions, in the cloud.

Azure functions provide data processing, system integration, working with the Internet of Things (IoT) and building simple APIs [5].

Google introduced the Android Things IoT platform with support for Intel Edison and Joule™ 570x microcomputers, NXP Pico i.MX6UL and Argon i.MX6UL, and Raspberry Pi 3 [6].

The company SONM has developed a multi-purpose decentralized platform for implementing complex computing tasks, based on blockchain and fog computing technologies [6].

Today, there are a wide variety of fog platforms, including private (Cisco IO, Nebbiolo Technologies, ClearBlade, Smartiply Fog, LoopEdge), public (Azure IoT, Amazon AWS IoT Greengrass, Google, Yandex and Mail.ru) and open source platforms. execution code (FogFrame2.0, FogFlow, FogBus) [7].

The key feature of fog computing technologies, which is to perform most of the data processing at the "edge" of the network, allows them to be used in a wide range of distributed systems, including monitoring and control systems, where system response time is one of the fundamental characteristics.

Monitoring and forecasting systems. In recent decades, the need to take into account information about the state of the environment has increased significantly. First of all, this is due to increased requirements for ensuring environmental safety, which, in turn, is the key to sustainable development of society. The intense impact of anthropogenic factors on the environment can cause various consequences, including negative natural phenomena that pose a danger to the population and various infrastructure facilities.

In this regard, it is relevant to use monitoring systems (MS) that ensure timely collection of data in order to notify the population about the expected occurrence of dangerous events [8]. The entire variety of currently existing monitoring systems can be divided into systems that do not process data, and systems that are capable of pre-processing data, thereby increasing the functionality of the system [9]. Currently, a promising direction for creating monitoring systems is the implementation of this class of systems based on "digital economy" technologies.

For example, Jonathan Bar-Magen Numhauser considers the problem of predicting natural hazards by processing large volumes of unstructured data generated by heterogeneous devices [1]. It is proposed to use fog computing technology to place some computing tasks on nodes of network equipment and mobile devices of organizations and individuals that are involved in monitoring. This technology allows you to reduce the load on the communication network and partially relieve the data processing center, which, in turn, increases the reliability of the system and reduces system latency.

The emergence of Internet of Things technologies and the development on their basis of "smart" cities, territories, enterprises and homes also contributes to the active implementation of monitoring systems in their infrastructure. Alex, er Slagg. [2] proposed a hierarchical distributed fog computing architecture to support the integration of a large number of infrastructure components and services in smart cities. As an example demonstrating the effectiveness of the proposed architecture, a prototype of an intelligent pipeline monitoring system in smart cities was implemented. Based on the results obtained, we can conclude that the use of fog technologies in the infrastructure of "smart" cities is promising, allowing them to significantly increase their "intelligence."

Erin Cunningham proposed an IoT-enabled intelligent water distribution and underground pipeline condition monitoring architecture for smart cities [3]. A key component of the developed architecture is the integration of Internet of Things technologies and fog and cloud concepts to effectively solve problems associated with automated water distribution, as well as monitor pipeline leaks.

Monitoring systems are actively used in agriculture. For example, Hamid Reza Arkian proposed an approach to data analysis and processing for distributed crop and soil monitoring, in which components that implement hierarchical data collection and modeling avoid problems associated with network bandwidth limitations and reduce the energy required for data transmission. Here the focus is on the use of fog computing technology, which allows the computing resources of local nodes to be used to further transfer partially processed data to higher levels of the system for decision making.

A model of a distributed library of a monitoring and diagnostic system for various areas of activity is also proposed (Fig. 3).

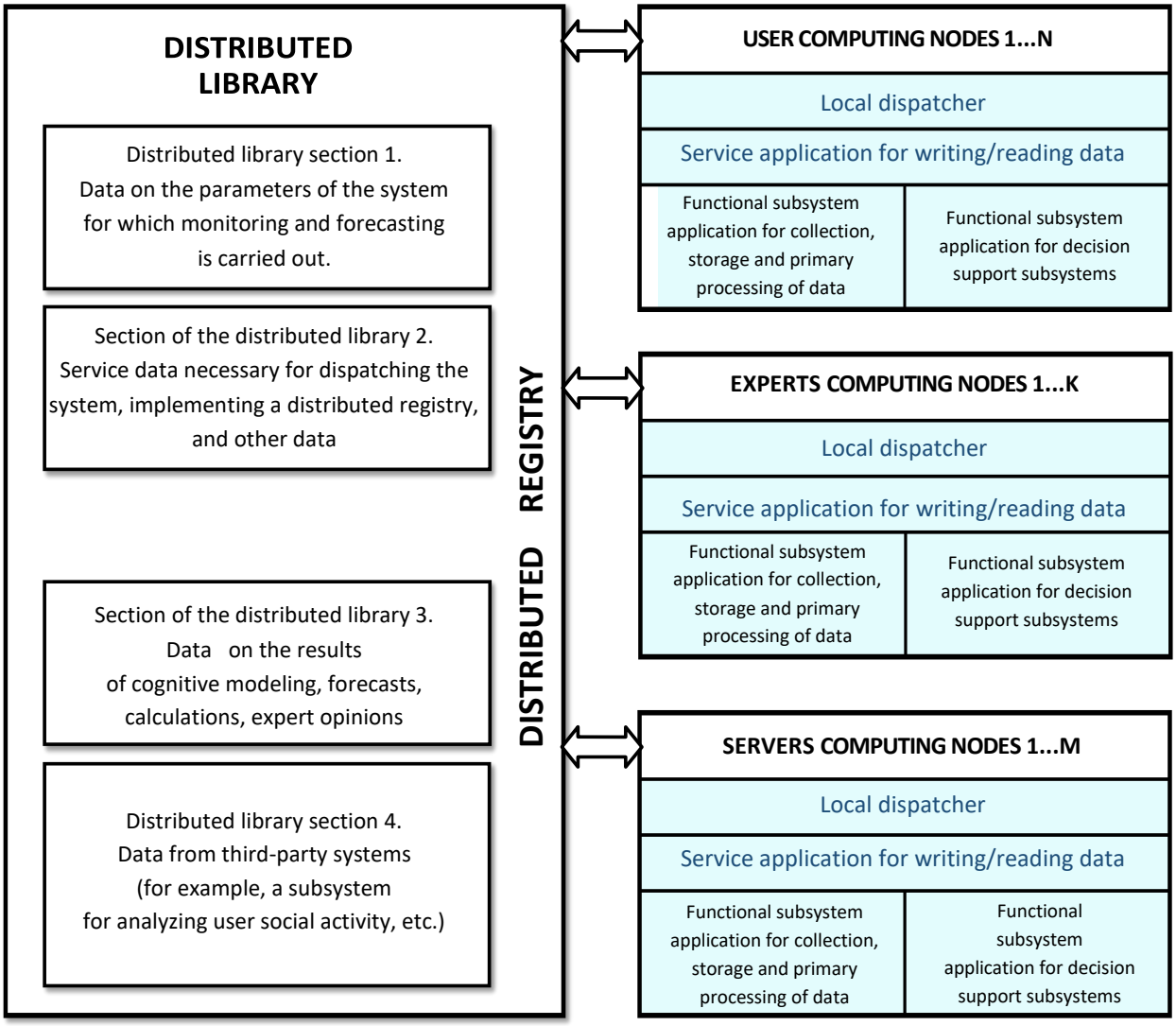


Fig. 3. Distributed library of monitoring and diagnostic system for various fields of activity

As can be seen from Fig. 3, the distributed library receives data from the data collection subsystem, their description, the results of their processing, after which they are available to users and experts. If necessary, the services provided by the subsystem can be executed on the cloud or fog computing layer. The decision support subsystem also has access to the distributed registry, and, therefore, to the data received from the collection subsystem. From this subsystem, data on the results of cognitive modeling and expert opinions can enter the distributed library. If necessary, simulation tasks can be executed on the computing nodes of the cloud and fog layers.

Conclusions

Thus, we can conclude that the construction of monitoring systems based on fog computing technologies is promising and effective in terms of reducing the load on the switching network, reducing system latency and partially unloading the data center for various areas of human activity, as evidenced by the above analysis.

REFERENCES

1. Rehan M.M., Rehmani M.H. Blockchain-enabled Fog and Edge Computing: Concepts, Architectures and Applications: Concepts, Architectures and Applications. Taylor and Fransis, 2020. 302 p.
2. Jonathan Bar-Magen Numhauser. Fog Computing- Introduction to a new Cloud evolution. Proceedings from the CIES III Congress, January 2012 (англ.) // Escrituras silenciadas: paisaje como historiografía / José Francisco Forniés Casals (ed. lit.), Paulina Numhauser (ed. lit.), Proceedings from the CIES III Congress, January 2012.
3. Hamid Reza Arkian, Abolfazl Diyanat, Atefe Pourkhalili. MIST: Fog-based data analytics scheme with cost-efficient resource provisioning for IoT crowdsensing applications // Journal of Network and Computer Applications. – 2017-03-15. – Vol. 82. – P. 152–165. – ISSN 1084-8045. doi: http://doi.org/10.1016/j.jnca.2017.01.012
4. Kuchuk G., Kovalenko A., Komari I.E., Svyrydov A., Kharchenko V. Improving big data centers energy efficiency: Traffic based model and method. Studies in Systems, Decision and Control, vol 171. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (Eds.). Springer Nature Switzerland AG, 2019. Pp. 161-183. DOI: http://doi.org/10.1007/978-3-030-00253-4_8
5. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об'єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: https://doi.org/10.20998/2522-9052.2018.1.04
6. Nechausov A., Mamusuĉ I., Kuchuk N. Synthesis of the air pollution level control system on the basis of hyperconvergent infrastructures. *Сучасні інформаційні системи*. 2017. Т. 1, № 2. С. 21 – 26. DOI: https://doi.org/10.20998/2522-9052.2017.2.04
7. Iervolino, R., Manfredi, S. Global stability of multi-agent systems with heterogeneous transmission and perception functions. Automatica. 2024. 162, 111510. DOI: http://doi.org/10.1016/j.automatica.2024.111510
8. Кучук Н. Г., Мерлак В. Ю., Скородєлов В. В. Метод зменшення часу доступу до слабкоструктурованих даних. *Сучасні інформаційні системи*. 2020. Т. 4, № 1. С. 97-102. doi: https://doi.org/10.20998/2522-9052.2020.1.14
9. Shi, H., Lin, W., Liu, C., Yu, J. A Novel Heterogeneous Parallel System Architecture Based EtherCAT Hard Real-Time Master in High Performance Control System. Electronics. 11(19), 3124. Doi: http://doi.org/10.3390/electronics11193124
10. She R., Sun M. Security Energy Efficiency Analysis of CR-NOMA Enabled IoT Systems for Edge-cloud Environment. *Int. Journal of Computational Intelligence Systems*. 2023. Vol. 16(1), 118. DOI: http://dx.doi.org/10.1007/s44196-023-00273-y.
11. Петровська І. Ю., Кучук Г. А. Розподіл обчислювальних ресурсів у хмарних системах. *Системи управління, навігації та зв'язку*. 2022. Вип. 2 (68). С. 75–78. DOI: http://dx.doi.org/10.26906/SUNZ.2022.2.075.
12. Kuchuk G., Nechausov S., Kharchenko, V. Two-stage optimization of resource allocation for hybrid cloud data store. *Int. Conf. on Information and Digital Technologies*. Zilina, 2015. P. 266-271. DOI: http://dx.doi.org/10.1109/DT.2015.7222982.
13. Кучук Г.А., Коваленко А. А., Лукова-Чуйко Н. В. Метод мінімізації середньої затримки пакетів у віртуальних з'єднаннях мережі підтримки хмарного сервісу. *Системи управління, навігації та зв'язку*. Полтава . ПНТУ, 2017. Вип. 2(42). С. 117-120.
14. Sharma, M., Kaur, P. Reliable federated learning in a cloud-fog-IoT environment. *Journal of Supercomputing*. 2023. Vol. 79(14). P. 15435–15458. DOI: http://dx.doi.org/10.1007/s11227-023-05252-w.
15. Baucas, M.J., Spachos, P. Improving Remote Patient Monitoring Systems Using a Fog-Based IoT Platform with Speech Recognition. 2023. *IEEE Sensors Journal*. Vol. 23(15). P. 17611–17618. DOI: http://dx.doi.org/10.1109/JSEN.2023.3287916.
16. Essalhi, S.E., Raiss El Fenni, M., Chafnaji, H. A new clustering-based optimised energy approach for fog-enabled IoT networks. *IET Networks*. Vol. 12(4). P.155–166. DOI: http://dx.doi.org/10.1049/ntw2.12082.

Received (Надійшла) 29.11.2023
Accepted for publication (Прийнята до друку) 24.01.2024

Технологія туманних обчислень в розподілених системах

З. В. Колесник, О. О. Меженський, О. П. Давикоза, Г. А. Кучук

Анотація. **Актуальність**. Концепція туманних обчислень є еволюційним етапом у розвитку концепції хмари. Вона займає провідне місце серед загальних тенденцій розвитку інформаційних технологій. Виникнення цієї концепції тісно пов'язане з виникненням і розвитком концепції Інтернету речей. **Результати.** Проаналізовано предметну область. Він включає в себе аналіз сучасних тенденцій в області організації розподілених обчислень, аналіз використання алгоритмів популяції та моделей онтології для вирішення оптимізаційних задач у розподілених системах, аналіз моделей, методів і алгоритмів для вирішення задачі передачі обчислювальних навантаження в розподілених системах, реалізованих на основі туманних обчислень. **Висновок.** Виявлено, що концепція туманних обчислень дозволяє вирішити більшість проблем, пов'язаних із навантаженням на комунікаційну інфраструктуру та затримкою обміну інформацією. Але вони не вирішують проблеми, пов'язані з високою динамічністю туманного середовища та супутнім зниженням ефективності розподіленої системи.

Ключові слова: комп'ютерна система, розподілена система, туманні обчислення, хмарні обчислення, Інтернет речей.

UDC 004.4 **doi: 10.26906/SUNZ.2024.1.098**

O. Kolesnikov, G. Golovko, V. Yastreba, Ye. Piatyntsev

National University «Yuri Kondratyuk Poltava Polytechnic», Poltava, Ukraine

LEVERAGING CLOUD TECHNOLOGIES AND SERVERLESS ARCHITECTURE FOR EFFICIENT WEB DEVELOPMENT: A CASE STUDY FROM REAL-WORLD APPLICATION

Abstract. This article provides a review of modern cloud solutions and serverless architecture using Backend as a Service (BaaS) and Function as a Service (FaaS) architecture as an example. In the scope of the article, a parallel is drawn between the consistently growing computing power and cloud technologies' growing popularity and availability for business. The results of an analytical review include a list of the most popular cloud providers from leading corporations and a comparison of low-level and high-level cloud technologies. The advantages and disadvantages of Google Cloud Platform (GCP) and Google Firebase are presented, where GCP is a low-level cloud provider and Firebase is a high-level cloud provider. The importance of understanding the context and specifics of the project when choosing cloud project solutions is emphasized which helps to maintain a balance between flexibility and development efficiency. The study introduces the practical utilization of the Supabase cloud platform for the development of a modern web application. The article convincingly proves the actuality of using Supabase for the development of an information system to optimize the modern personnel recruitment process, indicating specific advantages. An example of Supabase Edge Functions usage to generate feedback using the OpenAI Completions API and the Deno software platform is presented. The article convincingly proves that the use of cloud technologies is a modern strategy for building flexible, efficient, and scalable information systems. The advantages of the usage of the provision of infrastructure provided by world industry leaders are summarized.

Keywords: systems, information systems, information technologies, cloud providers, serverless architecture, GCP, Firebase, Supabase, efficiency, artificial intelligence9
.

Introduction

In today's world, due to the rapid development of technology, computing power is becoming cheaper and more powerful day by day. The growth of computing capabilities of supercomputers and servers is especially noticeable [1] (Fig 1).

In parallel with the growth of computing power, network technologies were also actively developed, which made it possible to significantly increase the speed of the Internet around the world, as well as make communication between hardware more stable.

One of the derivatives of the development of computing power and network technologies is the emergence of cloud technologies, which over the last decade have become an integral part of the best practices of modern software development.

The emergence of the concept of cloud computing marked a fundamental transformation of the concept of building the infrastructure of projects.

Organizations and companies have gained the ability to use significant computing and storage resources without the need for large initial investments in server infrastructure.

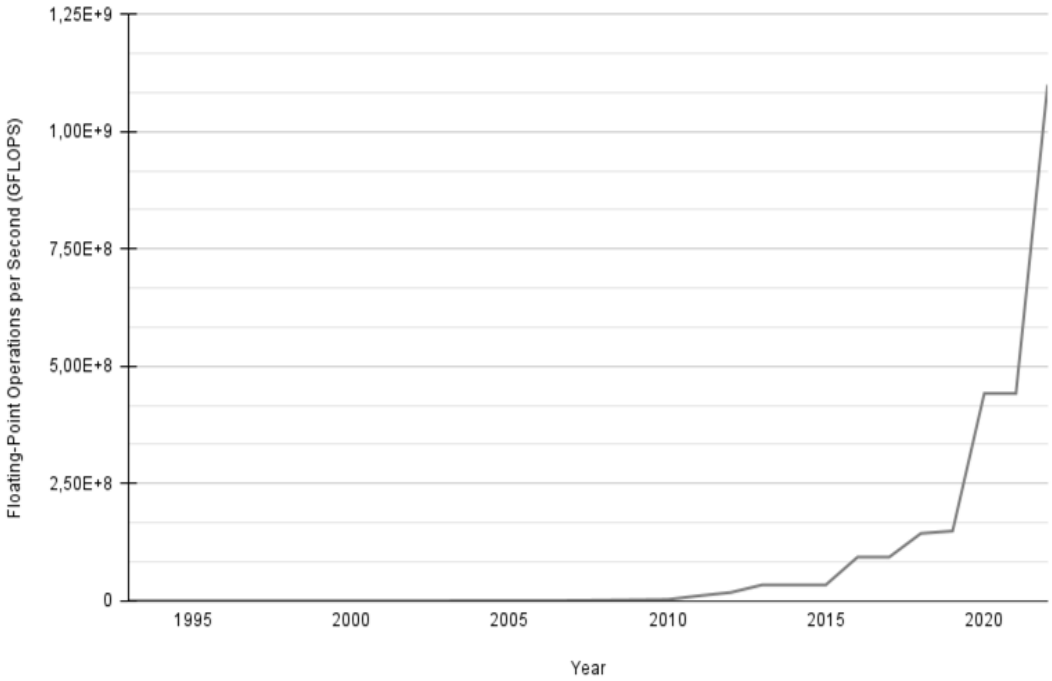


Fig. 1. Floating-point operations per second (GFLOPS)/year

Cloud technologies have gained particularly active use in web development, where serverless architecture has become one of the components for the effective and scalable development of web projects.

The introduction of serverless architecture has led to a move away from the use of traditional servers, which has opened up new opportunities for businesses and developers [2].

In the context of modern software development approaches, the concept of serverless architecture is often misunderstood.

It is commonly assumed that "serverless" means the complete absence of servers, but this is not true.

The proper meaning of the serverless architecture implies a change in the management and use of server resources.

In a traditional concept, teams are required to deal not only with the development of software code and business logic but also with the configuration and maintenance of the server infrastructure. This includes managing servers, ensuring their continuous operation, load balancing, and scaling resources according to the needs of the application.

However, a serverless architecture changes this approach by renouncing the need to directly manage servers. Instead, developers can focus solely on writing and deploying code using computing resources that are automatically allocated and scaled by the cloud provider. This provides efficiency, flexibility, and speed of deployment, reducing the burden on developers and allowing them to focus on creating an innovative product [2].

This article delves into the practical application of cloud technologies and serverless architecture in modern web development, using a real-world example that will provide valuable insight into the practical benefits and drawbacks of cloud technology usage.

The research purpose

The purpose of this article is to do a brief research on the existing cloud providers and solutions, analyze the pros and cons, and demonstrate a real-world example of serverless architecture usage in the information system to automate daily tasks in the recruitment workflow.

Analytical review of existing cloud solutions

The most famous cloud providers from global corporations are:

- Amazon Web Services (AWS);
- Google Cloud Platform (GCP);
- Microsoft Azure.

All three cloud providers are undisputed industry leaders that provide services that give confidence in the reliability of the infrastructure.

Cloud providers provide a wide variety of tools with different levels of abstraction. Depending on the level of abstraction, the ratio of flexibility to the speed of configuration and deployment changes. Lower-level tools are more flexible but more complex to configure, and high-level tools are less flexible but extremely easy and quick to configure which is ideal for Minimal Viable Product (MVP) or software development without complex server architecture requirements [2].

Google, in addition to the Google Cloud Platform, also provides higher-level cloud solutions named Firebase platform. Both Google Cloud Platform and Google Firebase have their advantages and disadvantages depending on the project. Firebase is a Backend as a Service (BaaS) platform that provides higher-level abstraction services that allow building mobile and web applications quickly and with minimal backend code and infrastructure setup.

For example, Firebase provides cloud database MongoDB, hosting, user auth, and other features [3].

Usage of low-level cloud platforms such as Google Cloud Platform include the advantages [3]:

1. Flexibility: GCP provides granular resource management, which is important for complex projects.
2. A wide range of services: from virtual machines (VMs) to network services and data storage.
3. Scalability: Suitable for large projects with complex architecture.

Disadvantages of low-level cloud platforms such as GCP are [3]:

1. Complexity of management: Requires more in-depth knowledge in the field of cloud resource management.
2. Higher costs: Some services may incur higher costs compared to more abstract platforms.

The advantages of high-level cloud platforms are [4]:

1. Simplified development: Firebase provides high-level tools that make it easy to build and manage applications.
2. Rapid deployment: Ideal for projects that require rapid development without complex infrastructure management.
3. Built-in features: Authentication, real-time databases, analytics, and more.

Considered disadvantages of high-level cloud platforms are [4]:

1. Limited flexibility: Firebase can be restrictive for very complex or specific project needs.
2. Platform Dependency: Focusing on using Firebase may lead to difficulties migrating to other platforms in the future.

The choice of platform always depends on the specifics of the project, but in the case of MVP project development, or a project that does not require complex architectural solutions, a platform like Firebase is an ideal solution, because, in addition to saving time, quite often a free quota is enough to test an idea or to make project working at the initial stage, which allows you to focus resources on the business logic of the project [3] (Fig. 2).

Cloud solutions are not ideal and might have different issues such as security or vendor-lock bottleneck, but at the same time, infrastructure is managed by industry leaders which helps to be sure that infrastructure is made following best practices.

The implementation and maintenance of in-house infrastructure require a significant amount of resources which might be critical for startups and small-grade projects [3].

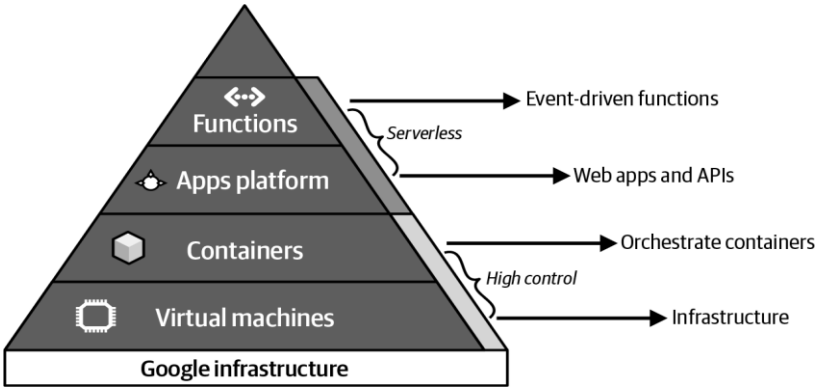


Fig. 2. Google infrastructure hierarchy

Main part

Based on the discovery made in the article "The objective need to implement an information system to automate daily tasks in recruitment workflow" the minimum necessary functionality for the implementation of the information system to automate daily tasks in the recruitment workflow is [7]:

- authentication;
- competency matrices management module;
- candidate management module;
- candidate interview module;
- a module for generating feedback for candidates based on interview results.

Based on the listed functionality, it is possible to conclude that the necessity:

- use of a modern DBMS for data storage, retrieval, and management;
- implementation of authentication or use of ready-made solutions;
- the presence of an API layer, which will avoid direct interaction of the client side with the database, which is considered anti-pattern.

The most effective way to implement a web application with the above-mentioned requirements is to use high-level Backend as a Service (BaaS) tools. BaaS is the serverless cloud model that helps the development team to deploy server-side logic without or with a minimal amount of coding. Usually, it might be a simple user interface [5].

The most popular BaaS solutions for web development are Firebase or Supabase. Supabase is a modern open-source platform that positions itself as an alternative to Google's Firebase. This platform provides a range of features and functionality that may be needed when creating modern web applications [8].

Architecturally Supabase is relative to the Backend as Service and Function as Service. FaaS is a serverless cloud model with a way to execute modular parts of code like specific functions on the edges. In specific cases like the development of the information system which should be proof of concept, Supabase is a platform tool that will help radically increase Time to Market (metric) [8].

One of the best advantages of the Supabase is its community-driven approach. It's open-sourced and is actively maintained by people across the world.

Also, Supabase has a much higher threshold of free quotas compared to Firebase which means that the product might live in the free tier for a longer time during growth. But even when the product is ready to scale the flexible pricing politics at the platform allow continuing growth without radical resource consumption increase.

Supabase provides a bunch of immediately ready-to-use solutions to save time during the development of the aforementioned application, especially the next ones [8]:

1. First things first Supabase deploys for free relational PostgreSQL which is manageable from the Supabase Dashboard with a lot of helpful interactive tools to manage the database. But the most important thing is that it still stays just a regular PostgreSQL database under the hood which might be used in the classic backend or to which it's possible to connect from any favorite software which means that we're not vendor-locked in Supabase.
2. Supabase Authentication provides ready-to-use free authentication functionality with various configurations, a flexible API for its use, and even ready-made UI elements. Various authentication methods are supported: starting with the classic method with a login and password and finishing with authentication with third-party providers such as Google, GitHub, and others.
3. One of the key advantages of Supabase is its authorization system, which uses PostgreSQL's row-level security (RLS) access control mechanism. This provides the ability to fine-tune accesses using specific conditions, thus allowing granular control of data access by different categories of users.
4. Edge Functions infrastructure which will help with OpenAI Completions API endpoint creation.
5. Supabase offers a convenient administration panel interface that allows you to manage services, database, and environment variables and perform monitoring, etc.
6. Supabase automates the process of APIs and related documentation creation using a database schema as a baseline. This provides convenience in interacting with the database through GraphQL or RESTful API endpoints.
7. Using Supabase Codegen to automatically generate TypeScript types ensures the accuracy of syncing database schemas with the client code, which simplifies development and reduces the risk of errors.

8. JavaScript SDK makes integration of the Supabase into the client-side code pretty quick and smooth. JavaScript SDK comes with predefined TypeScript types which in pair with Supabase codegen helps to avoid mistakes.

Supabase is not designed to resolve all problems by itself both with paid and free plans. For example, one of the biggest drawbacks of the Supabase usage on the free tier is the unavailability of database backup management from the user interface. But it's a feature that is pretty easy to do manually or even automate by CI/CD pipelines, for example using GitHub Actions. It's pretty important to choose solutions that do not lock the team from doing fallbacks. Supabase always gives the ability to use the fallback solutions where Supabase can't provide solutions from the box which makes this cloud provider a great choice from the scalability and maintenance point of view.

Features provided by Supabase are granular and useful in a standalone manner.

That means that each feature might be integrated or replaced down the road on demand which makes Supabase usage even more flexible (Fig 3).

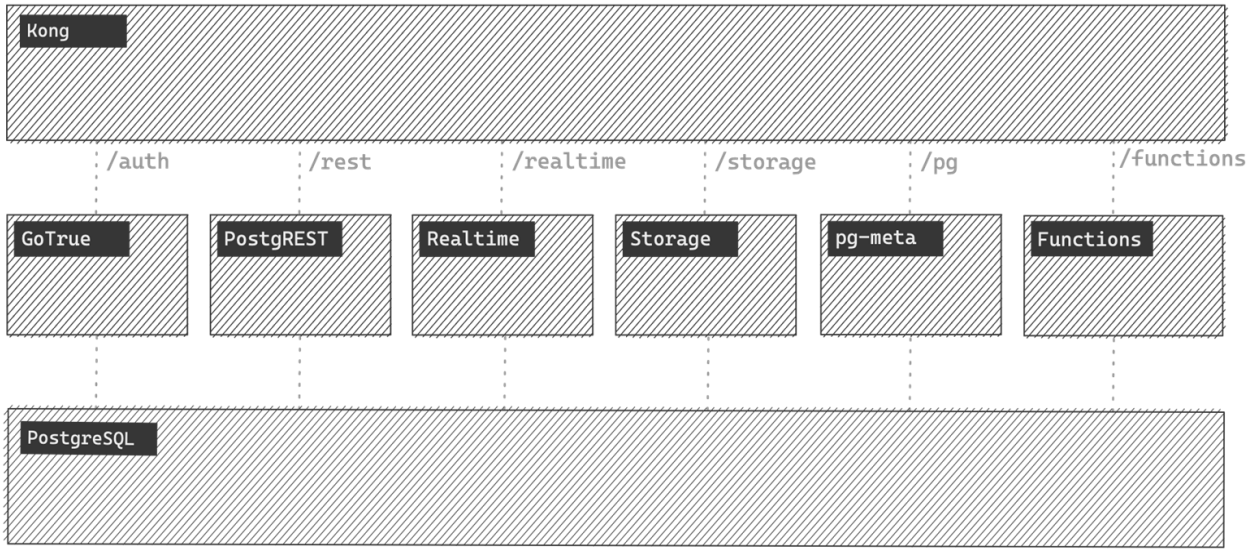


Fig. 3. Supabase architecture

Supabase Edge Functions and OpenAI Completions API. Edge functions are serverless JavaScript/TypeScript functions that are globally distributed across the world in a similar way to CDN. The main benefit of distributed globally serverless functions is reduced latency for final users of the information systems because of evaluation in data centers that are located as close to the user as possible. Edge functions might be used to achieve any common backend requirements. The reduced latency is critical in cases where performance is critical. Edge functions are a great fit for building automatic feedback generation using large language models. ChatGPT-like user experience is the best practice which gives users the most comfortable way of communication with LLM [6].

Edge functions are the implementation of the Function-as-a-Service (FaaS) and Serverless architecture patterns which means that those functions are deployable without the requirement to configure and manage server infrastructure which makes the deployment process easier [8].

Edge functions have next benefits [8]:

1. Increased efficiency of pricing because of the "pay as you go" pricing model. This pricing model is much more flexible compared to classic dedicated servers because the customer pays only for the resources that are used and pays nothing if the function is not used for a while.

2. Reduced latency for target users due to execution of the function in the data centers which is geographically closest to the requester.

3. Auto-scaling by design. Edge functions are designed in a way to make automatic scaling on demand. At the same time, scaling strategies are flexible because of the provided configuration API. For example, it's possible to always keep 1 or more instances running to avoid cold-start lag.

4. Edge functions are suitable for different use cases. They might be used for the classic API endpoint creation, dynamic content generation, webhooks, etc.

5. Edge functions provide an advanced security level thanks to its isolated nature. Independence from a centralized server environment radically decreases the chance of side effects from malicious requests for the whole system. Also, standalone nature helps to do maintenance more easily disabling only particular features of the application for the maintenance period.

6. For the specific use cases edge functions support Web Assembly, which makes them scalable and increases readiness for really unusual feature requests by businesses.

Supabase provides a big free quota and easy-to-use infrastructure for quick and reliable deployment of the edge functions based on the Deno platform which is a great fit for the development of the small startup or MVP applications. Also because of the usage of the Deno under the hood, it's possible to use a wide range of regular NPM

packages since Deno is the JavaScript/TypeScript platform. Deno was designed with a security-first and TypeScript-first vision which gives some advantages against Node.JS such as built-in TypeScript support without the need to transpile, better environment variables support, etc.

To use OpenAI Completions API for different use cases in the scope of the application, the edge function was designed abstracted from concrete features. This approach makes this function reusable and predictable. The Completions API settings were predefined in the const until they needed customization, but because of the clear API of the function, it remains ready for further enhancements on demand (Listing 1).

The authorization token is needed to make the OpenAI Completions API work. Deno's built-in environment variables tool was used since it's considered an anti-pattern to hard-code tokens in the codebase.

This helps to ensure that the token is not accidentally exposed to the codebase [6, 8].

```
const OPEN_AI_API_KEY = Deno.env.get("OPENAI_API_KEY");

const completionsApiSettings: CompletionsApiSettings = {
  model: "gpt-4",
  temperature: 0.7,
  top_p: 1,
  frequency_penalty: 0,
  presence_penalty: 0,
  max_tokens: 1000,
  stream: true,
  n: 1,
};

serve(async (req) => {
  const { messages } = await req.json();

  try {
    const res = await fetch("https://api.openai.com/v1/chat/completions", {
      method: "POST",
      headers: {
        Authorization: `Bearer ${OPEN_AI_API_KEY}`,
        "Content-Type": "application/json",
      },
      body: JSON.stringify({
        ...completionsApiSettings,
        messages: messages,
      }),
    });

    const processedCompletionsApiResponse = processCompletionsApiResponse(res);

    return processedCompletionsApiResponse;
  } catch (error) {
    return new Response(JSON.stringify({ data: null, error: error }));
  }
});
```

Listing 1. Code fragment with OpenAI Completions API usage in Supabase Edge Function

Conclusion

Most information systems require well-configured, secured, and scalable infrastructure. The classic approach to managing infrastructure gradually evolved into cloud-based infrastructure.

Keeping infrastructure in-house is expensive and requires time investment into the establishment and further maintenance. This article presented a chart of computing power growth over the years. This chart represents the amount of floating-point operations per second (GFLOPS)/year which is a standard way to estimate computing resources.

The provided chart proves that cloud technologies have become cheaper and more accessible to businesses.

An analytical review of existing cloud solutions presented in this article provides:

- list of the most popular cloud platform providers, compare;
- comparison of the low-level and high-level cloud solutions;

- compares the pros and cons of two cloud platforms from a single vendor (Google);
- highlights the importance of the project solutions decision-making based on project context and details.

In the main part was provided an example of the cloud platform selection for the implementation of the information system to automate daily tasks in the recruitment workflow based on defined requirements and features.

During the selection of the best-matching cloud solution selection, the level of abstraction of the cloud platform was determined.

A brief comparison of the most popular high-level cloud platforms such as Google Firebase and Supabase was made to make development efficient.

After a high-level comparison of the most popular Backend as a Service providers, pros and cons were analyzed to ensure that the open-source Supabase solution covers the project's requirements.

Also in the scope of the Supabase cons analysis, the fallback strategies possibility was estimated to ensure that Supabase doesn't make our application fully dependent and locked into the vendor.

In order to implement the feature of the feedback generation for candidates after interviews Supabase Edge Functions were utilized. Additionally, before implementation benefits of usage of the Supabase Edge Functions were highlighted. The fragment of the code with usage of the OpenAI Completions API is presented.

In conclusion, fast-growing computing performance leads to better pricing. Better pricing made a great environment for the growth of the new concept – cloud technologies and serverless architectures. Cloud providers in the last decades enhanced their infrastructure and nowadays customers who choose cloud solutions from top cloud providers might be sure of usage under the hood industry-standard infrastructure covered by best practices of top companies, high-security level, and automatic scalability by design. An ability to use cloud solutions is a big benefit for startups and small projects that can't hire experienced DevOps to manage it in-house.

References

1. Wong T. *Introduction to classical and quantum computing* / Thomas Wong. – [S. l.] : Rooted Grove, 2022.
2. Marinescu D. C. *Cloud computing: theory and practice* / Dan C. Marinescu. – [S. l.] : Elsevier Science & Technology, 2022.
3. Hunter T. *Google cloud platform for developers: build highly scalable cloud solutions with the power of google cloud platform* / Ted Hunter, Steven Porter. – [S. l.] : Packt Publishing, 2018. – 506 p.
4. Singh H. *Serverless Web Applications with React and Firebase: develop real-time applications for web and mobile platforms* / Harmeet Singh, Mayur Tanna. – [S. l.] : Packt Publishing, 2018. – 284 p.
5. Khan O. M. A. *Enterprise Application Architecture with .NET Core: An architectural journey into the Microsoft .NET open source platform* / Ovais Mehboob Ahmed Khan, Ganesan Senthilvel, Habib Ahmed Qureshi. – [S. l.] : Packt Publishing, 2017. – 564 p.
6. Sarrion E. *Exploring the Power of ChatGPT* [Electronic resource] / Eric Sarrion. – Berkeley, CA : Apress, 2023. – Mode of access: https://doi.org/10.1007/978-1-4842-9529-8 (date of access: 25.01.2024). – Title from screen.
7. Kolesnikov O. *The objective need to implement an information system to automate daily tasks in recruitment workflow* [Electronic resource] / O. Kolesnikov, G. Golovko // Системи управління, навігації та зв'язку. Збірник наукових праць. – 2023. – Vol. 3, no. 73. – P. 106–110. – Mode of access: https://doi.org/10.26906/sunz.2023.3.106 (date of access: 26.01.2024). – Title from screen.
8. Supabase documentation [Electronic resource] // Supabase Docs. – Mode of access: https://supabase.com/docs (date of access: 22.01.2024). – Title from screen.

Received (Надійшла) 22.11.2023
Accepted for publication (Прийнята до друку) 24.01.2024

Використання хмарних технологій та безсерверної архітектури для ефективної веб-розробки: приклад із реального світу

О. Колесніков, Г. Головко, В. Ястреба, Є. Пятінцев

Анотація. У цій статті розглянуто сучасні хмарні рішення та безсерверну архітектуру на прикладі використання Backend as a Service (BaaS) та Function as a Service (FaaS) архітектури. Проведено паралель між консистентно зростаючою обчислювальною потужністю та зростаючою популярністю і доступністю хмарних технологій. Здійснено аналітичний огляд в рамках якого наведено список найпопулярніших хмарних провайдерів від провідних корпорацій, порівняно високорівневі та низькорівневі хмарні технології. Представлено переваги та недоліки Google Cloud Platform (GCP) та Google Firebase, де GCP – низькорівневий хмарний провайдер, а Firebase – високорівневий. Підкреслено важливість розуміння контексту та особливостей проєкту при обранні хмарних проєктних рішень задля збереження балансу між гнучкістю та ефективністю розробки. Дослідження знайомить з практичним застосуванням хмарних технологій для розробки сучасного вебдодатку, а саме Supabase. Стаття переконливо доводить доцільність використання Supabase для рорзобки інформаційної системи для оптимізації сучасного процесу рекрутменту персоналу з зазначенням конкретних переваг. Представлено приклад використання Supabase Edge Functions для генерації зворотного зв`язку з використанням OpenAI Completions API та програмної платформи Deno. У статті переконливо доведено, що використання хмарних технологій є сучасною стратегією побудови гнучких, ефективних і масштабованих інформаційних систем. Узагальнено переваги використання інфраструктури, наданої світовими лідерами галузі.

Ключові слова: системи, інформаційні системи, інформаційні технології, хмарні провайдери, безсерверна архітектура, GCP, Firebase, Supabase, ефективність, штучний інтелект.

УДК 629.3 **doi: 10.26906/SUNZ.2024.1.104**

О. О. Копцев, В. О. Мартовицький, Н. М. Бологова, І. Б. Федак

Харківський національний університет радіоелектроніки, Харків, Україна

ОСОБЛИВОСТІ АВТОМАТИЧНОГО РОЗГОРТАННЯ ІНФРАСТРУКТУРИ ЯК КОДУ ДЛЯ ХМАРНИХ СЕРВІСІВ

Анотація. Хмарні сервіси надають сучасні обчислювальні ресурси, доступні на вимогу через Інтернет. Завдяки хмарним обчисленням команди стають більш ефективними та скорочують час виходу на ринок, оскільки вони можуть швидко набувати та масштабувати послуги без значних зусиль, які потребує управління традиційною інфраструктурою. Автоматизація дозволяє командам покращувати ключові показники. Команди відмовляються від тривалих процесів, пов'язаних із внесенням змін та запланованими розгортаннями. Вони також переходять від реактивного виявлення проблем до запобіжного моніторингу та забезпечення прозорості. Мета статті – дослідити популярні засоби для реалізації інфраструктури як коду, що включають Terraform, AWS CloudFormation, ARM Templates, Ansible, Puppet, Chef та інші. Ці інструменти допомагають створювати, керувати та відстежувати інфраструктурні ресурси через програмний код. Використання автоматизованих практик IaC дозволить зберегти час, зменшити ризики, підвищити сумісність та спростити процеси розгортання та управління інфраструктурою. Розглянувши популярні засоби для реалізації інфраструктури як коду, що допомагають створювати, керувати та відстежувати інфраструктурні ресурси через програмний код, ми дійшли висновку, що Bicep дозволяє більш ефективно та зрозуміло працювати з розгортанням інфраструктури в Azure, а також полегшує роботу з ARM Templates. Використання Bicep, у порівнянні з ARM шаблонами та іншими інструментами IaC, дає можливість створювати скрипти, які є значно компактнішими за розміром. Це досягається завдяки більш лаконічному та зрозумілому синтаксису Bicep, що дозволяє описувати однакові набори ресурсів меншою кількістю коду. Такий підхід не тільки спрощує розробку та підтримку інфраструктурного коду, але й знижує поріг входження для нових користувачів, які мають досвід роботи з програмуванням.

Ключові слова: хмарні сервіси, інфраструктура, масштабування, розгортання.

Вступ

Постановка проблеми. Хмарні сервіси (Cloud Services) - це послуги, які надаються через хмарну інфраструктуру, доступні через Інтернет [1]. Вони охоплюють широкий спектр обчислювальних та інших інформаційних послуг, які можуть бути використані організаціями та фізичними особами. Основною ідеєю хмарних сервісів є надання користувачам можливості використовувати обчислювальні ресурси, сховища даних та інші інфраструктурні ресурси за певну плату, без необхідності забезпечення та підтримки власного обладнання та програмного забезпечення.

Основні характеристики хмарних сервісів включають еластичність та масштабованість, що дозволяє користувачеві збільшити або зменшити обсяги використовуваних ресурсів залежно від потреб. Це забезпечує оптимальні обчислювальні потужності без зайвих витрат. Самообслуговування також є однією з основних характеристик хмарних сервісів. Користувачі можуть замовити ресурси та конфігурувати їх за допомогою веб-інтерфейсів або API, без потреби у взаємодії зі спеціалістами технічної підтримки. До характеристик хмарних сервісів також відносяться платіж за використання, доступність та відмовостійкість та широкий спектр послуг. Користувачі сплачують лише за той обсяг ресурсів, які вони зажадають. Це дозволяє зменшити витрати, втратити ви не платите за неактивний ресурс. Багато хмарних платформ забезпечують високу доступність за допомогою даних розташування та ресурсів на різних фізичних місцях. Це уникнути відмов у роботі у випадку збоїв обладнання. Хмарні сервіси можуть включати в себе обчислення, зберігання даних, аналітику, штучний інтелект, Інтернет-речі (IoT), розробки платформ та багато іншого.

Аналіз останніх досліджень і публікацій. Деякі провідні постачальники хмарних сервісів включають Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), IBM Cloud тощо. Ці послуги призначені для розподілених центрів обробки даних для забезпечення доступності та швидкості обробки. Моделі обслуговування хмарних сервісів забезпечують різні рівні гнучкості, контролю та відповідальності, що дозволяє користувачам вибирати найбільш підходящий варіант у залежності від їх потреб та навичок. Найбільш популярні [2]:

1. Інфраструктура як сервіс (IaaS - Infrastructure as a Service). У цій моделі користувачам надаються віртуальні обчислювальні ресурси, такі як віртуальні машини, мережеві ресурси та сховища даних. Користувач самостійно керує операційною системою, забезпеченням безпеки та розгортанням програмного забезпечення. Приклади: Amazon EC2, віртуальні машини Microsoft Azure.

2. Платформа як сервіс (PaaS - Platform as a Service). У цій моделі користувачам надається платформа для розробки, тестування та розгортання власних програмних додатків. Вони можуть працювати з програмами без необхідності в управлінській інфраструктурі. Платформа надає інструменти для розробки та керування додатками. Приклади: Heroku, Google App Engine.

3. Програмне забезпечення як сервіс (SaaS - Software as a Service). У цій моделі користувачам надається готове програмне забезпечення, яке вони можуть використовувати через Інтернет. Вони не контролюють над інфраструктурою або платформою, але підтримують можливість використання програми з

будь-яким пристроєм з доступом до Інтернету. Приклади: Google Workspace, Microsoft Office 365.

4. Функції як сервіс (FaaS - Functions as a Service) - це модель, у якій користувачам надаються можливості для розгортання окремих функцій або скриптів без необхідності керування обчислювальною інфраструктурою. Користувачі платять за фактичний час виконання функцій. Приклади: AWS Lambda, функції Azure.

5. Інфраструктура як код (Infrastructure as Code, IaC) - це модель автоматизованого управління та розгортання інфраструктурних ресурсів за допомогою програмного коду. Замість традиційного ручного налаштування та управління інфраструктурою, при використанні коду IaC для опису ресурсів, які повинні бути створені, налаштовані та керовані. Розгортання інфраструктури - це процес створення та налаштування деяких обчислювальних ресурсів, мережевих компонентів, сховищ даних та інших елементів інфраструктури для підтримки роботи програмних додатків. Цей процес може бути виконаний вручну, але часто він автоматизується за допомогою практики інфраструктури як коду (IaC) та відповідних інструментів [3]. Основні концепції та переваги IaC:

Автоматизація. IaC дозволяє автоматизувати процеси створення, налаштування та управління інфраструктурою. Це втратити ризики помилок та спростити процеси впровадження.

Версіонування. Код інфраструктури може бути збережений у системі контролю версій, що дозволяє відстежувати зміни, порівнювати версії та відновлювати попередні стани.

Гнучкість та швидкість. За допомогою IaC можна легко та безпечно розгортати нові середовища та змінювати конфігурацію.

Документація. Код інфраструктури служить як документація для самої інфраструктури. Він описує всі налаштування та компоненти системи.

Впровадження кращих практик. IaC дозволяє впроваджувати стандартизовані підходи та кращі практики до конфігурації та управління інфраструктурою.

Повторюваність. Код інфраструктури може бути легко переносимим між світовими середовищами, що дозволяє забезпечити узгодженість середовищ у різних етапах розробки та тестування.

Мета статті – дослідити популярні засоби для реалізації інфраструктури як коду, що включають Terraform, AWS CloudFormation, ARM Templates, Ansible, Puppet, Chef та інші. Ці інструменти допомагають створювати, керувати та відстежувати інфраструктурні ресурси через програмний код.

Виклад основного матеріалу

До основних кроків розгортання інфраструктури належать:

1. Визначення вимог. Спочатку потрібно чітко виконати вимоги до інфраструктури. Це може включати обчислювальні ресурси, сховища даних, мережеві налаштування, безпеку тощо.

2. Вибір інструментів. Треба вибрати підходящі інструменти та засоби для реалізації розгортання. Якщо використовується практика Інфраструктури як коду, обирають відповідний інструмент, наприклад, Terraform, CloudFormation, Ansible тощо.

3. Створення коду інфраструктури. Треба написати код, який описує всі необхідні ресурси та налаштування для інфраструктури. Цей код може бути структурованим за допомогою файлів або скриптів, залежно від обраного інструменту.

4. Тестування. Перед розгортанням конкуренції необхідно провести тестування коду інфраструктури на відповідність вимогам та наявним стандартам.

5. Розгортання. Процес розгортання, під час якого інструмент автоматично створює та налаштовує необхідні ресурси відповідно до вашого коду.

6. Конфігурація та налаштування. Після розгортання можна налаштувати, а також додатково налаштувати ресурси, які вже створені.

7. Моніторинг та управління. Після розгортання слід встановити моніторинг для відстеження працездатності інфраструктури, а також забезпечити систему управління, щоб реагувати на зміни та події.

Важливо пам'ятати, що розгортання інфраструктури - це динамічний процес, який може змінюватися з часом під час зміни вимог до програмного забезпечення. Традиційне розгортання, означає налаштування та встановлення інфраструктурних ресурсів, таких як сервери, мережеві компоненти, бази даних тощо, вручну або за допомогою засобів, які не є автоматизованими. Однак цей підхід стає менш популярним при застосуванні автоматизованого розгортання за допомогою практичної реалізації інфраструктури як коду (IaC). Основні етапи традиційного розгортання інфраструктури полягають у наступному:

Перший етап полягає у плануванні. Це визначення вимог до інфраструктури, вибір обладнання, обчислювальних ресурсів, мережевої архітектури тощо. Другий етап — це придбання фізичного обладнання або оренда серверів у дата-центрі. Наступний - фізична інсталяція. Розміщення серверів, налаштування мережі, підключення до джерел електроживлення, охолодження тощо. Четвертий етап полягає у встановленні операційної системи на кожному сервері, налаштування мережі, безпеки та ін. Наступний етап - встановлення та конфігурація сховищ даних (наприклад, база даних), які використовуються додатками. Далі - налаштування заходів безпеки, включаючи файрволі, антивіруси, моніторинг та інші заходи. Сьомий етап полягає у розгортанні програмного забезпечення: Встановлення та конфігурація програмного забезпечення, яке буде працювати на інфраструктурі. Восьмий етап — це тестування, виконання тестів для перевірки працездатності та стабільності системи. І останній етап — це підтримка та управління. Додаткове управління та підтримка інфраструктури, включаючи регулярне оновлення, моніторинг та відповідь на проблеми.

Цей традиційний підхід до розгортання інфраструктури може бути працездатним, але він може бути гнучким, вимагати більше часу на ручне налаштування та зміну інфраструктури. У сучасному світі більше підприємств переходять до автоматизо-

ваних практик IaC для забезпечення більшої ефективності та узгодженості в розгортанні та керуванні інфраструктурою. Автоматизовані практики інфраструктури як коду (IaC) - це підхід до розгортання та управління інфраструктурою, коли всі дії, пов'язані зі створенням, налаштуваннями та управлінням ресурсами, забезпечуються за допомогою програмного коду. Основною метою є автоматизація процесів, забезпечення сумісності, зменшення ризиків помилок та забезпечення гнучкості в управлінській інфраструктурі. Ключові практики IaC включають:

1. Декларативний код. Код IaC описує бажану структуру та налаштування інфраструктури, а не послідовність дій для його створення. Це дозволяє системі самостійно розпізнавати, які зміни потрібно зробити, для досягнення бажаного стану.

2. Керованість версій. Код IaC може бути збережений та відстежуваний у системі керування версіями, що дозволяє відновлювати попередні стани та вести спільну роботу в команді.

3. Масштабованість. За допомогою IaC можна легко розгортати та масштабувати ресурси відповідно до потреб.

4. Повторюваність. Код IaC може бути застосований для розгортання інфраструктури на різних етапах розробки та в різних середовищах (наприклад, розробка, тестування, продакшн).

5. Автоматизована та неперервна доставка. IaC допоможе забезпечити автоматичне розгортання та оновлення інфраструктури, що підтримує практику неперервної доставки.

6. Тести та перевірки. Код IaC може бути підданим автоматизованим тестуванням, щоб виявити помилки до розгортання.

Використання автоматизованих практик IaC дозволить зберегти час, зменшити ризики, підвищити сумісність та спростити процеси розгортання та управління інфраструктурою.

Доменно-орієнтована мова (Domain-Specific Language, DSL) при розгортанні Інфраструктури як коду (IaC) - це спеціалізована мова програмування або конфігурації, яка розроблена для виразного опису інфраструктури та її розгортання. Основна ідея перетворюється в тому, щоб мову, спеціально налаштовану для виразу концепцій і понять, що відповідає конкретному домену (у цьому випадку - розгортанню інфраструктури), замість використання загальної мови програмування. Відмінність DSL від загальних мов програмування, таких як Python, JavaScript чи Ruby, виникає в тому, що DSL спрощує виразність, зрозумілість та специфічність для конкретної області, в цьому випадку - управління інфраструктурою. Це робить код більш читабельним і зрозумілим для тих, хто працює в даній області.

Є кілька прикладів DSL для розгортання Інфраструктури як коду. Декларативний інструмент Terraform для створення, налаштування та управління інфраструктурою різних провайдерів (AWS, Azure, Google Cloud тощо) - це приклад DSL. Файли конфігурації Terraform містять декларативний код, який описує ресурси, їх взаємозв'язки та налаштування. Аналогічно, шаблони CloudFormation для AWS також є формою DSL, спеціалізованою на описі ресурсів та послуг Amazon Web Services.

Pulumi - це інша платформа, яка дозволяє писати IaC за допомогою загальних мов програмування, але з концепціями, призначеними для розгортання та управління інфраструктурою. Навіть Ansible, хоча і використовує мову Python, має спеціалізовану конструкцію для конфігурації інфраструктури та автоматизованого розгортання.

У Microsoft Azure використання доменно-орієнтованої мови (DSL) дозволяє описати і керувати різними аспектами інфраструктури та ресурсів, які розгортаються в хмарному середовищі Azure. Основним інструментом для використання DSL в Azure є мова ARM (Azure Resource Manager).

Шаблони Azure Resource Manager (ARM Templates) - це JSON-подібна мова, яка дозволяє створювати ARM-шаблони - описи файлів, які використовують ресурси, їх взаємозв'язки та налаштування. Це DSL для розгортання та управління ресурсами Azure. У шаблонах ARM ви можете описати різні ресурси - від віртуальних машин та мережевих складових до баз даних та інших послуг (лістинг 1).

```
{"$schema":
"https://schema.management.azure.com/schemas/20
19-04-01/deploymentTemplate.json#",
  "resources": [
    {
      "type": "Microsoft.Resources/resourceGroups",
      "apiVersion": "2020-10-01",
      "location": "eastus",
      "name": "myResourceGroup"
    },
    {
      "type": "Microsoft.Compute/virtualMachines",
      "apiVersion": "2019-07-01",
      "name": "myVM",
      "location": "[resourceGroup().location]",
      "properties": {
        "hardwareProfile": {
          "vmSize": "Standard_DS1_v2"
        },
        "storageProfile": {
          "imageReference": {
            "publisher": "MicrosoftWindowsServer",
            "offer": "WindowsServer",
            "sku": "2016-Datacenter",
            "version": "latest"}
        }
      }
    },
    {
      "type": "Microsoft.Storage/storageAccounts",
      "apiVersion": "2019-06-01",
      "name": "mystorageaccount",
      "location": "[resourceGroup().location]"
```

Лістинг 1

Azure CLI - це командний рядок Azure, який також має підтримку для використання DSL. Можна використовувати команди CLI для створення, налаштування та керування ресурсами за допомогою команд спеціалізованої мови (лістинг 2). Terraform можна використовувати з підтримкою Azure. Він також надає DSL для опису ресурсів та їх взаємозв'язків, незалежно від платформи хмарних послуг (лістинг 3).

```
az group create --name myResourceGroup --location
      eastus
az vm create \
  --resource-group myResourceGroup \
  --name myVM \
  --image Win2016Datacenter \
  --admin-username azureuser \
  --admin-password myPassword12
az storage account create \
  --name mystorageaccount \
  --resource-group myResourceGroup \
  --location eastus \
```

Лістинг 2

```
resource "azurerm_resource_group" "my_rg" {
  name     = "myResourceGroup"
  location = "eastus"
}
resource "azurerm_virtual_machine" "my_vm" {
  name                = "myVM"
  location            =
      azurerm_resource_group.my_rg.location
  resource_group_name   =
      azurerm_resource_group.my_rg.name
  network_interface_ids =
      [azurerm_network_interface.my_nic.id]
  vm_size             = "Standard_DS1_v2"
  os_profile {
    computer_name  = "myvm"
    admin_username = "azureuser"
    admin_password = "myPassword12"
  }
}
resource "azurerm_storage_account" "my_sa" {
  name                     = "mystorageaccount"
  resource_group_name      =
      azurerm_resource_group.my_rg.name
  account_tier             = "Standard"
  account_replication_type = "LRS"
}
```

Лістинг 3

Pulumi - це інструмент для інфраструктури як коду, який дозволяє використовувати звичайні мови програмування (наприклад, Python, JavaScript, Go) для опису інфраструктури. Він також підтримує Azure, що дозволяє використовувати DSL з мовами програмування для роботи з ресурсами Azure. У прикладі ми використовуємо мову програмування Python (лістинг 4).

```
resource_group =
      core.ResourceGroup('myResourceGroup')
vm = compute.WindowsVirtualMachine('myVM',
    resource_group_name=resource_group.name,
    admin_username='azureuser',
    admin_password='myPassword12',
    size='Standard_DS1_v2'
    storageProfile: {
      imageReference: {
        publisher: 'MicrosoftWindowsServer'
        offer: 'WindowsServer'
        sku: '2016-Datacenter'
        version: 'latest'
      }
)
storage_account =
      storage.Account('mystorageaccount',
    resource_group_name=resource_group.name,
    account_tier='Standard',
    account_replication_type='LRS',
)
```

Лістинг 4

Вісер - це мова декларативного опису інфраструктури, розроблена спеціально для Microsoft Azure. У файлі з розширенням .bicep визначається інфраструктура, яку можна розгорнути в Azure, а потім використовується цей файл протягом усього життєвого циклу розробки для повторного розгортання інфраструктури.

Основна ідея Вісер полягає в тому, щоб зробити опис інфраструктури більш зрозумілим, менш складним і більш структурованим. Вісер дозволяє більш ефективно та зрозуміло працювати з розгортанням інфраструктури в Azure, а також полегшує роботу з ARM Templates (лістинг 5).

```
resource myResourceGroup
   'Microsoft.Resources/resourceGroups@2020-10-01'
   = {
  name: 'myResourceGroup'
  location: 'eastus'
}
resource myVM
   'Microsoft.Compute/virtualMachines@2020-06-01' = {
  name: 'myVM'
  location: myResourceGroup.location
  properties: {
    hardwareProfile: {
      vmSize: 'Standard_DS1_v2'
    }
  }
}
resource myStorageAccount
   'Microsoft.Storage/storageAccounts@2020-08-01-
   preview' = {
  name: 'mystorageaccount'
  location: myResourceGroup.location
  sku: {
    name: 'Standard_LRS'
  }
  kind: 'StorageV2'
}
```

Лістинг 5

ARM (Azure Resource Manager) Templates і Вісер - це засоби для декларативного опису та розгортання інфраструктури в Microsoft Azure.

Порівняння двох цих засобів за критеріями: мова, зрозумілість, спостережуваність та модульність наведено у табл. 1.

Вісер дозволяє більш ефективно та зрозуміло працювати з розгортанням інфраструктури в Azure, а також полегшує роботу з ARM Templates. Однак, обираючи між ними, слід брати до уваги власні знання та потреби в конкретному проекті.

Висновки

Використання автоматизованих практик IaC дозволить зберегти час, зменшити ризики, підвищити сумісність та спростити процеси розгортання та управління інфраструктурою. Розглянувши популярні засоби для реалізації інфраструктури як коду, що допомагають створювати, керувати та відстежувати інфраструктурні ресурси через програмний код, ми дійшли висновку, що Вісер дозволяє більш ефективно та зрозуміло працювати з розгортанням інфраструктури в Azure, а також полегшує роботу з ARM Templates.

Таблиця 1 – **Порівняння засобів для декларативного опису та розгортання інфраструктури в Microsoft Azure**

Критерії порівняння до засобів	ARM Templates	Bicep
Мова	ARM Templates вибір JSON для опису ресурсів та їх налаштування. JSON є достатнім стандартним форматом, але великі шаблони можуть бути складними для читання та редагування.	Bicep використовує більш декларативний та структурований синтаксис, що надає властивості мови програмування. Це полегшує розуміння та редагування.
Зрозумілість	Доволі часто шаблони ARM можуть стати заплутаними через велику кількість JSON-коду та потребують вказати досить багато деталей.	Bicep робить код більш зрозумілим за допомогою свого синтаксису. Це особливо важливо при роботі з великими та складними шаблонами.
Спостережуваність	Якщо ARM Templates - це JSON, можна налагоджено відстежити зміни та розрізнити, які ресурси змінені чи додані.	Так як Bicep компілюється з шаблонами ARM, зміни стають більш видимими, щоб ви могли переглядати сирцевий Bicep-код, який згенерував певний шаблон ARM.
Модульність	Підтримка модульності та перевикористання коду не завжди є зручною, особливо для складних шаблонів.	Bicep має кращу підтримку модульності та перевикористання коду. Можна створювати власні модулі та використовувати їх для спрощення шаблонів.

Використання Bicep, у порівнянні з ARM шаблонами та іншими інструментами IaC, дає можливість створювати скрипти, які є значно компактнішими за розміром. Це досягається завдяки більш лаконічному та зрозумілому синтаксису Bicep, що дозволяє описувати однакові набори ресурсів меншою кількістю коду. Такий підхід не тільки спрощує розробку та підтримку інфраструктурного коду, але й знижує поріг входження для нових користувачів, які мають досвід роботи з програмуванням. Коротший і чистіший код в Bicep сприяє кращому розумінню та легшому відстеженню змін у процесах розгортання та управління інфраструктурою, роблячи Bicep більш ефективним інструментом у сфері IaC.

СПИСОК ЛІТЕРАТУРИ

1. Red Hat. What are cloud services? URL: https://www.redhat.com/en/topics/cloud-computing/what-are-cloud-services
2. Вікіпедія. Моделі обслуговування та існуючі рішення URL: https://uk.wikipedia.org/
3. What is Infrastructure as Code? [Електронний ресурс] // Mike Jacobs, Ed Kaim. – 2021. – Режим доступу: https://docs.microsoft.com/enus/devops/deliver/what-is-infrastructure-as-code
4. Guerriero M., Michele G. Adoption, support, and challenges of infrastructure-as-code: Insights from industry. In: 2019 IEEE international conference on software maintenance and evolution (ICSME). IEEE, 2019. p. 580-589.
5. Rahman O., Akond J., Rezvan A. A systematic mapping study of infrastructure as code research. Information and Software Technology, 2019, 108: p. 65-77.
6. Microsoft. Comparing JSON and Bicep for templates. URL: https://learn.microsoft.com/en-us/azure/azure-resource-manager/bicep/compare-template-syntax
7. Rahman O., Akond J. Gang of eight: A defect taxonomy for infrastructure as code scripts. In: Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering. 2020. p. 752-764.
8. Riti K. Pierluigi A. Infrastructure as Code. Beginning HCL Programming: Using Hashicorp Language for Automation and Configuration, 2021, p. 65-78.
9. Microsoft. What is Bicep? URL: https://learn.microsoft.com/en-us/azure/azure-resource-manager/bicep/overview?tabs=bicep
10. Bicep vs ARM Template URL: https://dev.to/evdbogaard/bicep-vs-arm-templates-bf9

Received (Надійшла) 22.11.2023
Accepted for publication (Прийнята до друку) 06.01.2024

Features of automatic deployment of infrastructure as code for cloud services

Oleg Koptsev, Vitalii Martovytskyi, Nataliia Bolohova, Ilko Fedak

Abstract. Cloud services provide modern computing resources available on demand over the Internet.Thanks to cloud computing, teams become more efficient and reduce time to market, as they can quickly acquire and scale services without significant efforts required for managing traditional infrastructure. Automation enables teams to improve key metrics. Teams get rid of lengthy processes associated with making changes and scheduled deployments. They are also moving from reactive problem detection to proactive monitoring and transparency. The goal of this article is to explore popular tools for implementing infrastructure as code (IaC), including Terraform, AWS CloudFormation, ARM Templates, Ansible, Puppet, Chef, and others. These tools help create, manage and monitor infrastructure resources through software code. Using automated IaC practices will save time, reduce risk, improve interoperability, and simplify infrastructure deployment and management processes. After looking at popular infrastructure-as-code tools that help create, manage, and monitor infrastructure resources through code, we came to the conclusion that Bicep allows you to work more efficiently and clearly with infrastructure deployment in Azure, and also makes it easier to work with ARM Templates. Using Bicep, compared to ARM templates and other IaC tools, makes it possible to create scripts that are much more compact in size. This is achieved thanks to the more concise and understandable syntax of Bicep, which allows describing the same sets of resources with less code. This approach not only simplifies the development and maintenance of infrastructure code, but also lowers the barrier to entry for new users with programming experience.

Keywords: Cloud services, infrastructure, scaling, deployment.

УДК 004.75, 004.318 **doi: 10.26906/SUNZ.2024.1.109**

В. А. Крилова, А. В. Івашко, О. О. Петренко

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

АНАЛІЗ ВАРІАБЕЛЬНОСТІ СЕРЦЕВОГО РИТМУ ЗА ДОПОМОГОЮ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

Анотація. У статті проведено стислий огляд та аналіз існуючих алгоритмів і програмних реалізацій діагностичних систем оцінки варіабельності серцевого ритму, що засновані на методах машинного навчання. Наведено переваги використання штучної нейронної мережі для класифікації типів електрокардіографічних сигналів, що забезпечує підвищення ефективності та якості функціональної діагностики серцевої діяльності. З метою виявлення найбільш ефективного варіанта побудови нейромережевих блоків для апаратно-програмного комплексу аналізу варіабельності серцевого ритму запропоновано декілька варіантів реалізації побудови нейронної мережі. Здійснено аналіз методів та алгоритмів морфологічного аналізу електрокардіограми та наведено основні етапи проектування штучної нейронної мережі у якості класифікатору розпізнавання образів - RR-інтервалів.

Ключові слова: нейрона мережа, машинне навчання, варіабельність серцевого ритму, діагностика серцевих захворювань, RR інтервал

Вступ

Використання нейронних мереж для виявлення, розпізнавання та класифікації об’єктів, зображень, сигналів в умовах сучасних цифрових технологіях знаходить все більш широке поширення в різних сферах науки і техніки. Прикладом таких задач є аналіз електрокардіограм поданих у цифровій формі, що відкриває нові можливості для оцінки стану людини за певними показниками, а також дозволяє проводити діагностичні досліди на новому рівні у реальному масштабі часу.

Великий відсоток захворювань людей пов’язаний із серцево-судинною системою. Тим часом, значну кількість проблем та ускладнень можна запобігти з допомогою неперервного моніторингу та постійного аналізу стану, зокрема знімаючи сигнали людського тіла та вчасно обробляючи їх. Одним із методів діагностики порушень серцево-судинної системи стосовно завдань профілактичної медицини є аналіз інформації про варіабельність серцевого ритму (ВСР) [1]. Оцінка ВСР заснована на математичному аналізі динаміки змін частоти серцевих скорочень. Оцінка діяльності серцево-судинної системи здійснюється шляхом реєстрації механічних, акустичних і біоелектричних проявів серцевої діяльності, найбільш доступних для реєстрації під час наркозу. Серед показників центральної та периферичної гемодинаміки найбільшу цінність представляють параметри серцевого ритму, артеріального і венозного тиску крові, серцевого викиду. Послідовний ряд кардіоінтервалів не є випадковим, а має складну структуру, що відображає важливі параметри серцево-судинної системи. Тому аналіз структури варіабельності серцевого ритму дає важливу інформацію щодо стану вегетативної регуляції серцево-судинної системи та організму в цілому. Зростаюча за законом Мура продуктивність сучасних обчислювальних систем та їх доступність відкрила нові можливості для обробки даних – використання раніше описаних математичних методів машинного навчання та систем штучного інтелекту. Такий підхід дозволив значною мірою виключити людину з процесу опрацювання інформації та здобувати нові дані з вже наявних.

Штучні нейронні мережі (ШНМ) є системами обробки інформації, які відрізняються від звичайних систем паралельним характером передачі інформації та наявністю процесу саморегуляції для забезпечення заданої цільової функції. Зазначені властивості сприяють їх застосуванню у медичній діагностиці електрокардіосигналу, який несе інформацію про варіабельність ритму серця.

Основною **метою** статті є аналіз та дослідження методів для кластеризації сигналів електрокардіограм та застосування ШНМ для вирішення завдання з аналізу та класифікації типів варіабельності серцевого ритму, що забезпечує підвищення ефективності та якості функціональної діагностики серцевої діяльності в цілому.

Аналіз існуючих та перспективних рішень. У різноманітних дослідженнях аналіз ВСР за допомогою нейронних мереж використовується для виявлення проблем у роботі серця. Запропоновано багато методів для класифікації ЕКГ, які різняться підходами, точністю, швидкістю опрацювання та іншими показниками. Класичні методи аналізу ЕКГ та проведення проб з навантаженням досягли на даний момент певної межи своїх діагностичних можливостей. У зв'язку з цим значний інтерес викликають нові методи високої роздільності для обробки ЕКГ, що дозволяють виділити та проаналізувати компоненти ЕКГ-сигналу, які несуть додаткову інформацію.

У роботі [2] нейронні мережі було використано для автоматичної класифікації п’яти класів аритмій серця (нормакардія, шлуночкова екстрасистолія, мерехтіння шлуночків, миготлива аритмія та блокада серця). Автори використовували лінійні та нелінійні методи аналізу ВСР для навчання нейронної мережі. Отримані результати показують, що запропоновані методи є ефективними для класифікації порушень серцевого ритму з прийнятною точністю. Поєднання лінійних та нелінійних функцій разом із використаним класифікатором підвищує ефективність процесу класифікації.

Спектральний аналіз ЕКГ, який реалізується обробкою сигналу за допомогою перетворення Фур'є, є одним з найбільш розповсюджених. Але значним недоліком такого методу є те, що частотні компоненти не можуть бути локалізовані у часі, що не дозволяє досліджувати динаміку змін частотних параметрів сигналу. При використанні цього методу, при збільшенні вікна обробки сигналу, відбувається покращення роздільності за частотою, але погіршується за часом, і навпаки. Таким чином неможливо визначити для певного моменту часу які спектральні компоненти присутні в сигналі. Для вирішення цих проблем розроблено метод вейвлет–перетворення нестаціонарних сигналів. У роботі [3] описується саме такий класифікатор. Автори розробили метод розпізнавання відхилень ЕКГ від норми на основі спільного застосування дискретного вейвлет-перетворення та ШНМ. Запропоновано метод аналізу та класифікації ЕКГ, що полягає у вейвлет-аналізі сигналів та нейромережевому розпізнаванні образів на основі багатошарового персептрону.

У статті [4] наведено методику класифікації сигналів з використанням нейронної мережі зустрічного розповсюдження та розглянуто моделі для класифікації випадкових та детермінованих сигналів. Також, у роботі було проведено дослідження залежності параметрів нейронної мережі від якості розпізнавання різних видів сигналів.

У роботі [5] представлені результати дослідження нейронних мереж для розпізнавання патологічних змін електричної активності серця. Проводилося порівняння багатошарового персептрону та модульної структури організації нейронної мережі. Автори сформували дві бази даних: патологій серця та аритмій. Дані були згруповані у кілька основних класів та поділені ще на дві незалежні частини: навчальну та тестову. Усі ваги нейромережі ініціалізуються випадковими величинами з рівномірним розподілом. Для виключення впливу випадковості на результати навчання всі дії повторюються по 3 рази для кожного значення кількості нейронів прихованого шару структури нейромережі яка досліджується. Далі проводиться багаторазове навчання ШНМ із різним обсягом прихованого шару. Дослідження показали, що підвищена чутливість до патології, низька похибка та можливість необмеженого розширення числа патологій які аналізуються, робить наведену модульну структуру, ймовірно, оптимальним вибором для вирішення задачі аналізу електрокардіосигналу.

У роботі [6] автор запропонував підхід до побудови нейронної мережі на основі багатошарового персептрона, що дозволяє провести більш точне розпізнавання. В роботі також проведено вибір відповідного алгоритму навчання, та показано, що для мережі типу багатошаровий персептрон забезпечується прийнятний результат при використанні методу сполучених градієнтів та класичного алгоритму зворотного розповсюдження помилки. В роботі наведено результати експериментів з перевірки працездатності та ефективності розробленої нейронної мережі на різних об'єктах природного походження.

Наведений огляд варіантів застосування ШНМ для вирішення різних науково-практичних завдань свідчить про їх широкі можливості та перспективність подальшого розвитку впровадження нейромережевих технологій там, де ставиться завдання: виявлення, розпізнавання та класифікації об'єктів, сигналів та зображень. Нейронні мережі активно вивчаються як інструмент для аналізу різноманітних біомедичних сигналів, зокрема – електрокардіограм. Виходячи з аналізу досліджень, можна зробити висновок що основним засобом аналізу ВСР з використанням машинного навчання є використання класифікаторів та багатошарових персептронів.

Методи дослідження

Як було вказано, ВСР є одним з найбільш опрацьованих та інформативних показників вегетативної активності. В основі багатьох систем скринінгу серця лежать алгоритми морфологічного аналізу ЕКГ: визначення положення R зубця і розташованих поруч Q- і S-зубців, які разом утворюють QRS комплекс (рис. 1).

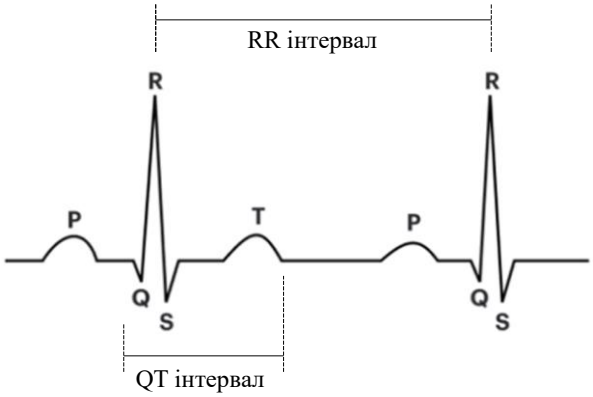


Рис. 1. RR інтервал на кардіограмі

Варіабельність серцевого ритму визначається як природні зміни інтервалів між серцевими скороченнями нормального ритму серця [7], які називаються RR (за позначенням R зубців) інтервалами. Іншими словами, це ступінь коливань тривалості інтервалів між синусовими комплексами, які зумовлені впливами відділів вегетативної нервової системи, а також гуморальними чинниками, навколо середнього рівня.

Дослідження варіабельності ґрунтується на вимірюванні інтервалів часу між R зубцями електрокардіограм та побудови на їх основі ритмограми з подальшим її аналізом за допомогою різних математичних методів. Непостійність інтервалу між кардіоциклами знаходиться у межах деякої середньої величини, що є оптимальною для обраного функціонального стану організму.

Через це варіабельність визначають у статичних станах організму, бо при будь-якій їх зміні, частота серцевих скорочень починає теж змінюватись, підлаштовуючись під новий функціональний рівень – виникає перехідний процес, під час якого починають працювати системи, що не пов’язані з регулюванням ВСР організму.

Основні методи аналізу варіабельності серцевого ритму [7]:

1. **Статистичний метод** засновано на аналізі змін RR-інтервалів, а також на порівнянні показників, що дають кількісну оцінку варіабельності. При їх використанні кардіоінтервалограма роздивляється як сукупність послідовних проміжків часу – інтервалів RR. Статистичні дані включають SDNN, SDANN, RMSSD, pNN5O, СКО (середньоквадратичне відхилення) та CV (коефіцієнт варіації).

2. **Геометричний метод** відображає розподіл RR – інтервалів. Метод використовує варіаційну пульсограму, за довжиною якої відкладаються значення RR – інтервалів, а по висоті – частота потрібних кардіоінтервалів (рис. 2). Суть геометричного методу (побудова гістограми) полягає у визначенні закону розподілу кардіоінтервалів як випадкових величин. При цьому будується крива розподілу кардіоінтервалів (гістограма) та визначаються її основні характеристики:

- мода – значення кардіоінтервалу що зустрічається найбільш часто у даному динамічному ряді;
- варіаційний обсяг X – різниця між максимальним та мінімальним значенням RR – інтервалів. Відображає ступінь варіативності значень кардіоінтервалів у динамічному ряду що вивчається;
- тріангуляційний індекс варіабельності серцевого ритму – загальна кількість RR-інтервалів, поділених на висоту гістограми всіх RR-інтервалів. Цей показник малочутливий до різного роду помилок, що виникають при поділі комплексів QRS на нормальні і ненормальні.

Тим самим знижуються вимоги до якості запису ЕКГ і її аналізу.

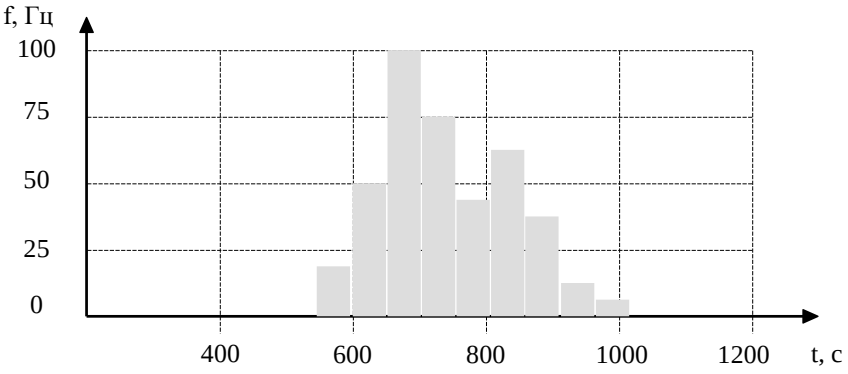


Рис. 2. Гістограма RR-інтервалів

3. **Спектральний аналіз ритму серця.** В основі методу лежить хвильова мінливість серцевого ритму, яка аналізується за допомогою перетворення Фур'є, що розбиває загальний спектр та її складові діапазони хвиль (табл. 1). Аналіз спектральної щільності потужності коливань дає інформацію щодо розподілу потужності в залежності від частоти коливань. Застосування спектрального аналізу дозволяє кількісно визначити різноманітні частотні складові коливань ритму серця та наочно графічно представити співвідношення різних компонентів серцевого ритму, що відображають активність визначених ланок регуляторного механізму. Виділяють три основних спектральних компонента, які відповідають коливанням ритму серця різної періодичності:

- High frequency (HF) – високочастотний компонент спектра, підвищується у стані спокою, під час сну, при гіпервентиляції, знижується – при фізичному навантаженні, стресі, різноманітних захворюваннях серцево-судинної системи;
- Low frequency (LF) – низькочастотний компонент спектра, у пацієнтів з тяжкою серцевою недостатністю виражена симпатична активація поєднується з істотним зниженням потужності LF;
- Very low frequency (VLF) – потужність хвиль дуже низької частоти, відображає активність повільно діючих гуморальних механізмів регуляції серцевого ритму, які зумовлюють закономірні зміни протягом тривалих проміжків часу (дні, тижні);
- Ultra Low Frequency (ULF) – ультра низькочастотні, які використовують для тривалих записів.

Таблиця 1 – **Компоненти спектру**

Назва компоненту спектру	Частотний діапазон, Гц	Період, сек
HF	0,4-0,15	2,5-6,6
LF	0,15-0,04	6,6-25,0
VLF	0,04-0,015	25,0-66,0
ULF	Менше 0,015	Більше 66,0

4. **Метод кореляційної ритмограми** полягає у графічному відображенні послідовних пар кардіоінтервалів (попереднього та наступного) у двомірній координатній площині. При цьому по осі абсцис відкладається велична $R - R_n$, а по осі ординат – величина $R - R_{n+1}$. Графік та область точок, які отримано таким чином, є кореляційною ритмограмою або скатерограмою (scatter-розсіювання) (рис. 3).

Таким чином, метод кореляційної ритмографії більш компактно відображає лінію кардіоінтервалів, та незалежно від часу проведення дослідження – хвилини або години. Цей спосіб оцінки ВСР відноситься до методів нелінійного аналізу та є корисним для випадків, коли на фоні монотонності ритму зустрічаються рідкі та раптові порушення (ектопічні скорочення та (або) «випадіння» окремих серцевих скорочень).

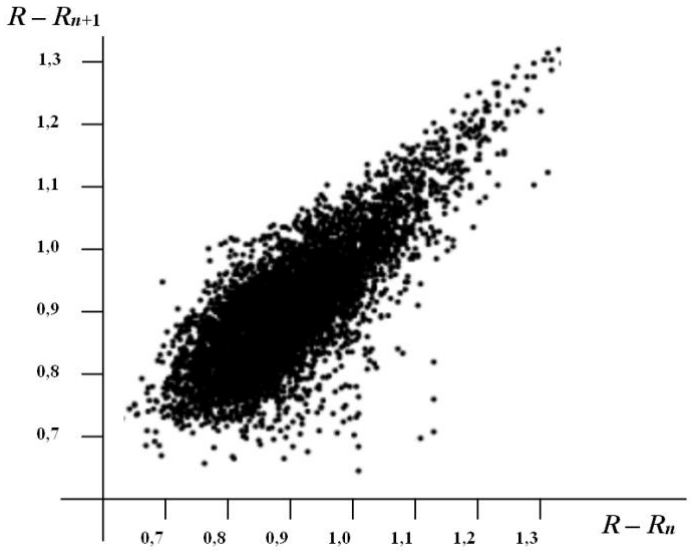


Рис. 3. Скатерограма RR інтервалів – нормальний синусоїдальний ритм

Геометричні методи дозволяють оцінити фізіологічний стан людини з точки зору розпізнавання аритмій, коли методи статистичного, і спектрального аналізу варіабельності серцевого ритму малоінформативні або неприйнятні, в цьому випадку доцільно використовувати оцінку кореляційної ритмограми. Аналіз гістограми і скатерограми ритму серця є більш коректним для оцінки нестаціонарних процесів, які характерні для біологічних систем.

Особливості реалізації ШНМ для аналізу ВСР

Для побудови та проектування автоматизованої системи діагностики пацієнтів з серцево-судинними захворюваннями та оцінки ефективності застосування ШНМ для аналізу варіабельності серцевого ритму необхідно вирішити наступні питання.

1. Вибір та обґрунтування архітектури нейронної мережі. С початку задається кількість прихованих, вхідних та вихідних шарів та синаптичні зв'язки між нейронами. Як зазначалося вище, модель багатошарового персептрону є найбільш широко вивченою, він складається з одного вхідного та одного вихідного шару, з одним або декількома прихованими шарами [8]. Таким чином, стандартна багатошарова нейронна мережа складається з трьох шарів вузлів, які з'єднані між собою за допомогою синаптичних ваг ω_{ij} та ω_{ik}, як зображено на рис. 4.

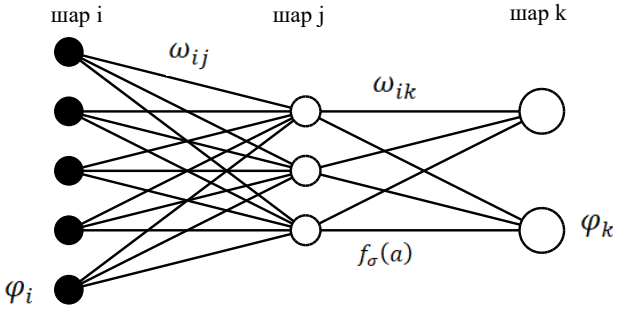


Рис. 4. Схема 5-3-2 багатошарової нейронної мережі

Вхідні блоки передають вхідні дані, так само як нелінійні вихідні одиниці кінцевого рівня отримують дані від кожного з блоків у прихованому шарі. Одиниці зміщення підключаються безпосередньо через вагові коефіцієнти зміщення до кожного з нейронів у прихованих і вихідних шарах. Топологія міжнейроних синаптичних зв'язків задається на етапі проектування мережі, але під час навчання та тестування системи, якщо потрібно, архітектура мережі може корегуватися.

2. Вибір функції активації нейрона мережі $f_\sigma(a)$ та алгоритмів навчання ШНМ. В якості активаційної функції кожного окремого нейрона була обрана безперервна сигмоїдальна біполярна функція [9]

$$f_\sigma(a) = \frac{1}{1+e^{-a}}. \quad (1)$$

Цю функцію пропонується використовувати для синтезу ШНМ для аналізу ВСР.

Вихідне значення нелінійного нейрону отримується за формулою:

$$\varphi = f_\sigma\left\{\left(\sum_i \omega_i x_i\right)\right\}. \quad (2)$$

Після того як вихід нейронів одного шару під'єднано до входу іншого шару - буде сформована єдина нейронна мережа.

Але, оскільки аналіз ВСР відноситься до задачі класифікації сигналів, метою яких є визначення належності певного об'єкта до відповідного класу, доцільним є використання методів та алгоритмів навчання з вчителем.

Для цього необхідно задати відповідні ваги ω_{ij} та ω_{ik} для синаптичних зв'язків, які в процесі навчання мережі будуть корегуватися в залежності від відповіді ШНМ та навчальних значень.

Вхідні дані, що використовуються для навчання мережі (далі позначені як φ_i) подаються до мережі та поширюються скрізь неї, щоб отримати результат φ_k:

$$\varphi_k = f_\sigma(\sum_j \omega_{jk} f_\sigma(\sum_i \omega_{ij}\varphi_i)). \quad (3)$$

Під час навчання цільові дані або бажаний результат t_k, який асоціюється з навчальними даними, порівнюються з фактичним результатом φ_k. Після чого коригуються ваги ω_{jk} та ω_{ij}, для того щоб мінімізувати різницю між фактичним та цільовим значенням [10]. Ця різниця (похибка) визначається протягом усього навчання шаблонів p для навчального набору як

$$E = \frac{1}{2}\sum_p \sum_k \left(f_\sigma\left(\sum_j \omega_{jk} f_\sigma\left(\sum_i \omega_{ij}\varphi_i^p\right)\right) - t_k^p \right)^2. \quad (4)$$

Квадрат похибки E може бути мінімізованим за допомогою методу градієнтного спуску. Відповідний градієнт розраховується щодо кожної ваги ω_{ij} та ω_{jk}. Рівняння оновлення ваги для прихованих та вихідних шарів наступні:

$$\omega_{jk}^{(\tau+1)} = \omega_{jk}^{\tau} - \eta \frac{\delta E}{\delta \omega_{jk}}, \quad (5)$$

$$\omega_{ij}^{(\tau+1)} = \omega_{ij}^{\tau} - \eta \frac{\delta E}{\delta \omega_{ij}}, \quad (6)$$

де η – ступінь швидкості навчання.

Середньоквадратична похибка між кожним вхідним і вихідним шаблоном, підсумована за всіма шаблонами, є критерієм, який використовується під час навчання.

3. Формування тестових та навчальних вибірок RR-інтервалів. Для тренування та тестування нейронної мережі було використано набір даних MIT-BIH Arrhythmia Database придатний для завдань обробки сигналів ЕКГ. Згідно з описом баз даних на ресурсі PhysioNet, MIT-BIH містить півгодинні уривки двоканальних амбулаторних записів ЕКГ, отриманих від суб'єктів [11]. Записи оцифровані з частотою дискретизації 360 відліків на секунду на канал з 11-бітною роздільною здатністю (діапазон 10 мВ). Окрім власне кардіограм, база містить мітки класів типів серцевого ритму. В цілому, для навчання моделі нейронної мережі доцільно використовувати набір не менше ніж з 10 тисяч коротких (на більше 5 хв.) одноканальних записів ЕКГ пацієнтів з нормальним ритмом та різними відхиленнями. Для того щоб сформувати ці фрагменти, було обрано випадковим чином двадцять три записи із набору 24 годинних амбулаторних записів ЕКГ, які зібрані у змішаної популяції стаціонарних та амбулаторних пацієнтів, 60% та 40% відповідно. Ще двадцять п'ять записів, були відібрані з того самого набору, щоб включити менш поширені, але клінічно значущі аритмії, які не були добре представлені в попередній випадковій вибірці.

4. Обробка ЕКГ та створення масиву нормованих RR-інтервалів. До того, як дані потрапляють на вхід нейронної мережі, вони проходять додаткову обробку і первинний аналіз. При аналізі ЕКГ, в першу чергу, важливим є періодичний QRS комплекс, а саме R-зубець комплексу, є початком пульсової хвилі, тому вони дуже важливі в процесі аналізу ЕКГ, адже можна окремо визначати окремі хвилі та аналізувати їх. Оскільки QRS комплекс легко визначити за R зубцем, вікно у 0,5 секунди з обох боків зубця використовується для визначення початку кожного серцебиття та ВСР можна визначати як інтервали між індексами QRS комплексів. Існує багато алгоритмів виявлення R зубця у QRS комплексі з ЕКГ, з них найбільш поширеними є алгоритм Пана і Томпкінса. Так як вхідні образи (значення R зубця) являють собою вектори різної довжини, необхідно виконати нормалізацію сигналів для приведення амплітуд у встановлений динамічний діапазон для вхідних значень нейронної мережі. Створення масиву нормованих RR-інтервалів відбувається з використанням нелінійної функції виду:

$$\tilde{x}_i = f\left(\frac{x_i - \tilde{x}_i}{\sigma_i}\right);\ f(a) = \frac{1}{1+e^{-a}}, \quad (7)$$

де $\tilde{x}_i = \frac{1}{p}\sum_{a=1}^{p} x_i^a$ –

середнє значення RR інтервалів;

$$\sigma_i^2 = \frac{1}{p-1}\sum_{a=1}^{p}(x_i^a - \tilde{x}_i)^2 \ -$$

дисперсія.

Після такого перетворення значення RR-інтервалів розподіляються за законом, близьким до рівномірного, що теоретично має покращувати якість навчання нейронної мережі.

Далі значення RR-інтервалів підлягають математичному аналізу для отримання статистичних параметрів [12]:

- середнє значення RR-інтервалів;
- середнє значення серцевого ритму;
- стандартне відхилення величин нормальних інтервалів N-N протягом 24 год;
- стандартне відхилення різниці послідовних інтервалів N-N;
- відсоток послідовних інтервалів N-N, різниця між якими перевищує 50 мс.

Вихідним значенням методу є масив з розрахованими параметрами.

5. Визначення вихідних даних ШНМ для оцінки ВСР. Вихідними даними є графіки кардіограм з позначеними R-зубцями, графік зміни довжини RR-інтервалів, а також текстові дані зі значеннями варіабельності та результатами її класифікації відповідно до відомих класів серцевого ритму та вірогідності приналежності до цього класу.

Для більш якісного навчання моделі нейронної мережі, вона потребує налаштування додаткових параметрів, які визначають також швидкість навчання, оцінку якості навчання, оптимізацію визначення та зсуву ваг нейронів. Окрім параметрів, які визначають як нейронна мережа буде працювати взагалі, потрібно також визначити, як вона буде обробляти дані, які надано для навчання. Щоб отримати якісь суттєві значення ваг нейронів, потрібно

кілька разів пройти навчання на тренувальному наборі даних (епохи). Мала кількість епох не дозволить якісно навчити персептрон, велика – потребує великих втрат часу та обчислювальних ресурсів, а також призводить до перенавчання - ситуації коли мережа більше завчає тренувальні дані (разом з шумами), ніж фактично навчається. Для навчання персептрону рекомендується 50 епох – подальше навчання не призводить до суттєвих змін у точності отриманої моделі.

Висновки

Нейромережевий аналіз має достатню гнучкість, забезпечує нелінійну обробку вихідних даних, має хорошу узагальнюючу здатність та можливість гнучкого навчання. Перевага застосування наведеного (модульного) варіанту нейронної мережі полягає в концентрації ресурсів кожного модуля на розпізнаванні лише одного класу, що має сприяти зменшенню ймовірності помилки та невірного висновку для всієї системи загалом. Крім того, модульний варіант ШНМ системи дозволяє розширювати функціональні можливості шляхом збільшення кількості доступних для аналізу ВСР модулів без перенавчання всієї системи. Застосування існуючих методик знаходження оптимальної кількості нейронів у прихованому шарі для розглянутих структур ШНМ на основі забезпечення рівномірного розподілу значень чутливості, специфічності та точності по кожному класу, сприяють підвищенню ефективності роботи нейронної мережі та дають можливість вибору такого поєднання значень чутливості та точності, при яких забезпечується максимальна достовірність розпізнавання ВСР.

З метою пошуку оптимальних параметрів навчання необхідно провести додаткові експериментальні дослідження та порівняльний аналіз багатошарового персептрону та структури модульного варіанту ШНМ як можливих варіантів побудови системи нейромережевих блоків аналізу ВСР.

СПИСОК ЛІТЕРАТУРИ

1. Clifford G.D., Liu Ch., Moody B., Lehman L.H., Silva I., Li Q. Classification from a Short Single Lead ECG Recording: The PhysioNet – Computing in Cardiology Challenge 2017, “Computing in Cardiology”, pp.1-4, DOI: 10.22489/CinC.2017.065-469.
2. Чернетченко Д.В., Мілих М.М., Луданов К.В. Апаратна реалізація імпульсної імпульсної штучної нейронної мережі для детектування параметрів електрокардіографічного сигналу (ЕКГ). Дніпропетровський національний університет ім. Олеся Гончара, Том 4, 2019 (275). DOI: 10.31891/2307-5732-2019-275-4-126-133
3. Tekeste T., Saleh H., Mohammad B., Khandoker A., Elnaggar M. A nano-watt ecg feature extraction engine in 65nm technology, IEEE Trans. on Circuits and Systems II: Express Briefs PP (99) (2017) 1–1. DOI: 10.1109/TCSII. 2017.2658670.
4. Jain S., Ahirwal M., Kumar A., Bajaj V., Singh G. QRS detection using adaptive filters: A comparative study, ISA Transactions 66 (2017) 362–375. DOI: 10.1016/j.isatra.2016.09.023.
5. Лісун Ю.Б., Углев Є.І. Варіабельність серцевого ритму, використання та методи аналізу. ДНУ «Центр інноваційних медичних технологій НАН України». № 4. 2020. DOI: 10.25284/2519-2078.4(93).2020.220693.
6. Karimipour M., Homaeinezhad R. Real-time electrocardiogram p-qrs-t detection delineation algorithm based on quality-supported analysis of characteristic templates, Computers in Biology and Medicine 52 (2014) 153–165. DOI: 10.1016/j.compbiomed.2014.07.002.
7. Kovalchuk M., Kharchenko V., Yavorskyi A. ECG signal classification using machine learning techniques. Bulletin of Taras Shevchenko National University of Kyiv. Series Physics & Mathematics 2022, 2 DOI: 10.17721/1812-5409.2022/2.9
8. Wu L., Xie X., Wang Y. (2021): ECG Enhancement and R-Peak Detection Based on Window Variability, “Healthcare” 2021, (Basel), 9 (2), P. 227; DOI: 10.3390/healthcare9020227.
9. Wieclaw L., Khoma Y., Fałat P., Sabodashko D., Herasymenko V. Biometrie identification from raw ECG signal using deep learning techniques. In 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). Vol. 1. P. 129–133). DOI: 10.1109/IDAACS.2017.8095063
10. Goldberger A. L., Amaral A. N., Glass L. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research for complex physiologic signals. Circulation. 2000 Vol. 101. P. 215-220.
11. PhysioNet Content Overview URL: https://physionet.org/about/content/
12. Bailey J.A. et.al. Behavioral simulation and synthesis of biological neuron systems using synthesizable VHDL, Neurocomputing, Elsevier B.V., pp. 2392-2406, 2011. DOI: 10.1109/BMAS.2008.4751231.
13. Дудикевич В.Б., Хома В.В., Чекурін В.Ф., Хома Ю.В. Нормалізація сигналів ЕКГ для застосування в системах біометричної ідентифікації Інформатика, обчислювальна техніка та автоматизація. Том 30 (69) Ч. 1 № 4 2019. DOI: 10.32838/2663-5941/2019.4-1/10.

Received (Надійшла) 12.12.2023
Accepted for publication (Прийнята до друку) 07.02.2024

Heart rate variability analysis using artificial neural networks

Viktoriia Krylova, Andrey Ivashko, Oleh Petrenko

Abstract. The article provides a brief review and analysis of existing algorithms and software implementations of diagnostic systems for assessing heart rate variability, based on machine learning methods. The advantages of using an artificial neural network to classify types of electrocardiographic signals are presented, which improves the efficiency and quality of functional diagnostics of cardiac activity. To identify the most effective option for constructing neural network blocks for the hardware-software complex for analyzing heart rate variability, several options for implementing the construction of a neural network have been proposed. An analysis of methods and algorithms for morphological analysis of an electrocardiogram is carried out and the main stages of designing an artificial neural network as a classifier for pattern recognition such as RR intervals are given.

Keywords: neuron, machine learning, heart rate variability, diagnosis of heart disease, RR interval.

UDC 004.7 **doi: 10.26906/SUNZ.2024.1.115**

Andrii Kuliahin

National Aerospace University named after M.E. Zhukovsky "KHAI", Kharkiv, Ukraine

PERSONALIZATION OF VISUAL CONTENT OF INTERACTIVE ART IN AUGMENTED REALITY BASED ON INDIVIDUAL USER PREFERENCES

Abstract. Topicality. In connection with the development of AR technologies and their use in interactive art, there is a growing need to develop methods of personalizing visual content, focused on the individual preferences of users. **Research methods.** Neural collaborative filtering method, generalized matrix factorization method, mood analysis on video. **The purpose of the article**: Researching the possibilities of improving the personalization of visual content in interactive art by evaluating the emotional reactions of users and their implicit feedback. **The results obtained.** The application of neural collaborative filtering and generalized matrix factorization to create adapted visual content in interactive art in AR was considered, which will significantly increase the relevance and immersion of users in interactive works. **Conclusion.** The considered approach can be used to improve immersiveness and personalization during user interaction with interactive art in AR.

Keywords: interactive art, augmented reality, neural collaborative filtering, generalized matrix factorization.

Introduction

The development of augmented reality (AR) technologies and their application in interactive art [1] opens new opportunities for personalization of visual content. Personalization becomes the basis for creating a deeper and more meaningful experience for users, allowing art to adapt to individual preferences and emotional states [2]. However, there are often problems with improving the immersion effect during user interaction with interactive art in augmented reality systems.

This article focuses on the novelty of the content personalization approach through the use of recommender systems based on implicit user feedback and the analysis of the user's mood expressed in the video (Video Sentiments Analysis) to improve the quality of personalization.

One of the aspects of improving the effect of immersion in art systems is the adaptation of content and interface to the needs and preferences of the user. In particular, the use of implicit feedback to the user can contribute to increasing the immersion effect and simplifying the user interface. Meanwhile, oversaturation of the system with explicit feedback can reduce the level of immersion and overload the interface [3, 4].

Implementation of a comprehensive approach to the personalization of visual content in interactive art, based on the application of Neural Collaborative Filtering (NCF) and Generalized Matrix Factorization (GMF) for processing user feedbacks. This approach allows taking into account not only the explicit choices and preferences of users, but also implicit signals such as emotional reactions and behavioral patterns during interaction with art objects. This depth of analysis is a significant advance in the field of personalized interaction, as it paves the way for creating unique, emotionally resonant experiences for each user.

In addition, the innovativeness of the research is emphasized by the use of Video Sentiments Analysis algorithms for a detailed study of the emotional reactions of users to visual content. This allows not only to improve the accuracy of the system's recommendations, but also to provide a deeper understanding of the emotional impact of art on an individual. Thus, the research makes a significant contribution to the development of personalized technologies in the field of augmented reality, opening new perspectives for interactive art.

The purpose of the article: the idea of this article is a detailed analysis and discussion of methods of personalization of visual content in interactive art in AR, with an emphasis on the use of implicit user feedback and Video Sentiments Analysis (VSA).

We aim to show how these approaches can significantly improve the user experience, making artistic experiences more personalized.

Recommendation systems in art personalization

Recommender systems have become a cornerstone of content personalization, revolutionizing the way users interact with content in the digital age. The latest trends are the use of recommender systems to personalize the art experience. Such systems use advanced algorithms and data analytics to curate art collections according to individual tastes, thereby increasing user engagement and satisfaction [5]. At the heart of this transformation is the ability of recommender systems to analyze vast amounts of data, including user interactions, preferences and reviews, to predict which artworks a user might like.

In the field of art personalization, recommender systems use different techniques such as collaborative filtering, content-based filtering, and hybrid models to offer personalized art experiences. Collaborative filtering analyzes patterns of past user behavior to recommend artworks that similar users like, while content-based filtering recommends art based on the characteristics of artworks the user has expressed interest in. Hybrid models combine both approaches, refining recommendations to provide a more nuanced and personalized art discovery experience [6, 7].

The authors of the publication «Hybrid Recommendations and Dynamic Authoring for AR Knowledge Capture and Re-Use in Diagnosis Applications» [8] explore the potential of hybrid recommendation systems

and dynamic authoring in the context of augmented reality for knowledge capture and its reuse in diagnostic applications. They point to the expansion of the capabilities of recommender systems beyond their traditional uses, emphasizing integration with complex technological solutions to improve diagnostic applications.

The impact of recommender systems on art personalization goes beyond simple convenience. This not only broadens the audience for artists and galleries, but also fosters a deeper cultural appreciation among the public.

In addition, these systems offer artists and curators valuable insights into audience preferences, guiding them to create and curate art that resonates more deeply with audiences. As recommendation technologies continue to evolve, their integration into art platforms promises to further enrich the landscape of art consumption, making art more accessible, engaging and personal than ever before.

NCF and GMF for processing user feedbacks

The use of neural collaborative filtering (NCF) and generalized matrix factorization (GMF) to process user feedback is becoming a best practice in the field of personalized recommendations. NCF, which uses deep neural networks to model interactions between users and objects, enables the detection of complex nonlinear dependencies in interaction data. This greatly improves the accuracy of recommendations, as NCF is able to gain a deeper understanding of users' subtle preferences and interests that may not be apparent using traditional methods [9].

On the other hand, GMF extends the classical factorization matrix approach by integrating it with neural networks to improve big data processing and discover hidden interaction factors. GMF uses a linear combination of latent characteristics of users and objects, offering a more accurate representation of their relationships. This allows recommender systems to not only more accurately predict potential user interest, but also take into account a wider range of user behaviors and feedback, ensuring high relevance of recommended content.

As the study conducted in the work «Study of methods for building recommendation system to solve the problem of selecting the most relevant video when creating virtual art compositions» [10] showed, the most effective approach to solving the problem of building a recommendation system of virtual art compositions is the approach hybridization, which consists in the combination within one model of different methods of building recommender systems, namely, the collaborative filtering method, the content-based method, and the knowledge-based method. The hybrid model, which combines all three methods, showed better results compared to models that implement each method separately. This is due to the fact that an additional deep neural network added to the hybrid of matrix factorization and collaborative filtering methods takes into account user characteristics when determining the video rating in the virtual art composition.

So, by combining NCF and GMF, a strong recommender system can be built that efficiently processes user feedback, providing personalized suggestions that take into account both explicit and implicit user preferences. This helps create a deeper and more engaging user experience, increasing user satisfaction and loyalty to the service. Therefore, the integration of NCF and GMF opens new horizons for the development of personalized services that can adapt to unpredictable changes in user preferences and behavior.

Using VSA to increase personalization

Let's analyze the latest publications related to VSA. Some current research focuses on sentiment analysis in YouTube comments. For example, in the article «Sentiment Analysis on Online Videos by Time-Sync Comments. Entropy» [11] authors use the methods of sentiment analysis and topic clustering to study educational content on YouTube. In particular, the authors consider how these techniques can be used to analyze comments to determine popular themes, emotional connection to material, and the overall effectiveness of educational content. The article presents different approaches to sentiment analysis, such as machine learning and deep learning, and their application to identify positive, negative, or neutral emotions in the textual content of comments.

Authors of publication «Learning Analytics on YouTube Educational Videos: Exploring Sentiment Analysis Methods and Topic Clustering» [12] describes the use of sentiment analysis for time-synchronized comments on videos. The authors focus on identifying and analyzing the emotional reactions of viewers at specific moments of the video, which allows for a dynamic understanding of content perception. The research findings show that this approach can be used to improve engagement and optimize video content, and provide clues for video content editors about how video affects the emotional state of viewers.

Using Video Sentiment Analysis (VSA) to improve personalization is an advanced approach that uses artificial intelligence to decode human emotions from video content. This innovative method opens up new ways to tailor user experience, content recommendations, and interactive services by understanding and responding to users' emotional states. VSA technology processes the video input by identifying facial expressions, body language and voice tones to determine the viewer's feelings in real time. This analysis provides a deep understanding of how content affects emotions, which can be used to personalize experiences in ways that resonate more deeply with each individual [13, 14].

Thus, the article «A Closer Look at Spatiotemporal Convolutions for Action Recognition» considers the possibility of using 3D CNN for multimodal spatiotemporal motion recognition [15]. This model can be adapted and used for VSA. In this way, it will be possible to achieve immediate user feedback for a recommender system working with interactive art.

Using recommender systems together with VSA to personalize interactive art opens up exciting prospects for creating unique art experiences. This combination allows not only to tailor artwork to the user's individual preferences based on their previous interactions, but also

to respond to their emotional state in real time, taking into account feedback from emotion analysis. This approach not only increases the individualization of the art consumption experience, but also creates a deeper emotional connection between the user and the art object, making interaction with art more meaningful and immersive.

Conclusion

The use of recommender systems based on processing implicit user feedback and video sentiment analysis in interactive art in AR opens up new perspectives for personalization of visual content. This approach allows not only to increase the satisfaction of users from works of art, but also to deepen the emotional connection between the work and the consumer. The introduction of these technologies has the potential to radically transform the way the user interacts with modern interactive art, making each experience unique and unique.

It is advisable to conduct further research with the inclusion of the generation of new interactive art content for each user, leveling or limiting the number of manually created virtual art compositions.

REFERENCES

1. Gironacci, Irene. (2021). State of the Art of Extended Reality Tools and Applications in Business. 10.4018/978-1-7998-4339-9.ch008.
2. Chen, Rongfei & Zhou, Wenju & li, yang & Zhou, Huiyu. (2022). Video-Based Cross-Modal Auxiliary Network for Multimodal Sentiment Analysis. IEEE Transactions on Circuits and Systems for Video Technology. PP. 1-1. 10.1109/TCSVT.2022.3197420
3. Wang, Fei. (2023). Research on the application of immersive art in digital technology scene. Advances in Education, Humanities and Social Science Research. 5. 88. 10.56028/aehssr.5.1.88.2023.
4. Zhang, Ying. (2023). Immersive Multimedia Art Design Based on Deep Learning Intelligent VR Technology. Wireless Communications and Mobile Computing. 2023. 1-8. 10.1155/2023/9266522.
5. Li, Huihong. (2023). Personalized Art Work Recommendation System and Methods Based on User Interest Characteristics and Emotional Preferences. Scalable Computing: Practice and Experience. 24. 883-894. 10.12694/scpe.v24i4.2393.
6. Patel, Dhruval & Patel, Foram & Chauhan, Uttam. (2023). Recommendation Systems: Types, Applications, and Challenges. 2210-142. 10.12785/ijcds/130168.
7. Duraisamy, Premkumar & Natarajan, Yuvaraj & .S, Yuvaraj & V.Niranjani,. (2023). An Overview of Different Types of Recommendations Systems - A Survey. 10.1109/ICITIIT57246.2023.10068631.
8. Fernández del Amo Blanco, Iñigo & Erkoyuncu, John & Farsi, Maryam & Ariansyah, Dedy. (2021). Hybrid recommendations and dynamic authoring for AR knowledge capture and re-use in diagnosis applications. Knowledge-Based Systems. 239. 107954. 10.1016/j.knosys.2021.107954.
9. He, Xiangnan & Liao, Lizi & Zhang, Hanwang. (2017). Neural Collaborative Filtering. Proceedings of the 26th International Conference on World Wide Web.
10. Kuliahin, Andrii & Narozhnyi, V. & Tkachov, V. & Kuchuk, H.. (2022). ДОСЛІДЖЕННЯ МЕТОДІВ ПОБУДОВИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧІ ВИБОРУ НАЙБІЛЬШ РЕЛЕВАНТНОГО ВІДЕО ПРИ СТВОРЕННІ ВІРТУАЛЬНИХ АРТ-КОМПОЗИЦІЙ. Системи управління, навігації та зв'язку. Збірник наукових праць. 4. 94-99. 10.26906/SUNZ.2022.4.094.
11. Li, Jiangfeng & Li, Ziyu & Ma, Xiaofeng & Zhao, Qinpei & Zhang, Chenxi & Yu, Gang. (2023). Sentiment Analysis on Online Videos by Time-Sync Comments. Entropy. 25. 1016. 10.3390/e25071016.
12. Chalkias, Ilias & Tzafilkou, Katerina & Karapiperis, Dimitrios & Tjortjis, Christos. (2023). Learning Analytics on YouTube Educational Videos: Exploring Sentiment Analysis Methods and Topic Clustering. Electronics. 12. 3949. 10.3390/electronics12183949.
13. Li, Jiangfeng & Li, Ziyu & Ma, Xiaofeng & Zhao, Qinpei & Zhang, Chenxi & Yu, Gang. (2023). Sentiment Analysis on Online Videos by Time-Sync Comments. Entropy. 25. 1016. 10.3390/e25071016.
14. Deshmukh, Rushali & Amati, Vaishnavi & Bhamare, Anagha & Jadhav, Aditya. (2023). Visual Sentiment Analysis: An Analysis of Emotions in Video and Audio. 10.1007/978-981-99-6586-1_21.
15. Tran, Du & Wang, Heng & Torresani, Lorenzo & Ray, Jamie & LeCun, Yann & Paluri, Manohar. (2017). A Closer Look at Spatiotemporal Convolutions for Action Recognition.

Received (Надійшла) 24.11.2023
Accepted for publication (Прийнята до друку) 07.02.2024

Персоналізація візуального контенту інтерактивного мистецтва в доповненій реальності на основі індивідуальних уподобань користувачів

А. І. Кулягін

Анотація. Актуальність. У зв'язку з розвитком технологій AR та їх використанням у інтерактивному мистецтві, зростає потреба в розробці методів персоналізації візуального контенту, орієнтованих на індивідуальні вподобання користувачів. **Методи дослідження.** Метод нейронної колаборативної фільтрації, метод узагальненої матричної факторизації, аналіз настрою на відео. **Мета статті:** Дослідження можливостей покращення персоналізації візуального контенту в інтерактивному мистецтві через оцінку емоційних реакцій користувачів та їх неявних відгуків. **Отримані результати.** Було розглянуто застосування нейронної колаборативної фільтрації та узагальненої матричної факторизації для створення адаптованого візуального контенту в інтерактивному мистецтві в AR, що дозволить значно підвищити релевантність та зануреність користувачів в інтерактивні твори. **Висновок.** Розглянутий підхід може бути використаний для покращення імерсивності та персоналізації під час взаємодії користувача з інтерактивним мистецтвом в AR.

Ключові слова: інтерактивне мистецтво, доповнена реальність, нейронна колаборативна фільтрація, узагальнена матрична факторизація.

УДК 519.7004.3 **doi: 10.26906/SUNZ.2024.1.118**

О. Г. Лебедєв, О. В. Бондар, Є. О. Самойленко, В. Г. Черевко

Харківський національний університет радіоелектроніки, Харків, Україна

АНАЛІЗ ІСНУЮЧИХ ПІДХОДІВ ДО РОЗРАХУНКУ КІЛЬКІСНОЇ ОЦІНКИ ЖИВУЧОСТІ DRONES

Анотація. На сьогоднішній день живучість відображає стійкість до диструктивних впливів як окремих підсистем drone, так і drone загалом. Така живучість закладена в алгоритмічну частину підсистем drone і дозволяє у разі виникнення нештатних ситуацій змінювати послідовність роботи підсистем drone. При вирішенні задачі аналізу та синтезу елементів підсистем drone, оперують кількісною оцінкою живучості. Це необхідно для визначення, з якою ймовірністю відмови окремих елементів будь-якої підсистеми drone призведуть до пошкодження drone. В даний час серед розробників drones не існує єдиної думки для визначення кількісної оцінки живучості drones. Дуже багато залежить від архітектури підсистем drone, їх взаємодії між собою та взаємодії їх елементів між собою. Ця стаття присвячена дослідженню підходів до розрахунку кількісної оцінки живучості drones на етапі проектування.

Ключові слова: «'Swarm-bot» - system, «s-bot», живучість, безпека, надійність, невразливість, стійкість, реінжиніринг, комунікація, nano-drone.

Вступ

Постановка проблеми. Останнім часом з'явився термін drone, що означає відсутність пілота на його борту (рис. 1). Це вимагає від drone певної автономності.

Рис. 1. Приклад drone

Фундаментальним у концепції автономії є розуміння відмінностей у базових термінах: автоматичні drones; автономні drones [1]. Автоматичні drones вміють самостійно досягати поставленої мети, дотримуючись запрограмованої логіки. Автономні drones спроможні самостійно впоратися з позаштатними ситуаціями, підключаючи заздалегідь запрограмований набір правил, який допомагає їм зробити правильний вибір. Прийнято виділяти чотири базові рівні автономії для drones [2].

Перший рівень автономії - контрольований та керований оператором drone. У такому drone оператор приймає всі рішення щодо функціонування drone. Такий drone не здійснює автономного контролю physical environment. Другий рівень автономії – контрольований, але не керований оператором drone. Такий drone може виконувати поставлені перед ним завдання, коли оператор дає drone певні дозволи. У такому режимі drone може ініціювати певні дії на основі даних, одержаних від бортових сенсорів. Ініціація може відбуватися лише у межах виконання поточного завдання. Третій рівень автономії – повноваження оператора делеговані drone. Такий drone може виконувати поставлені завдання, які не залежать від контролю оператором. При такому режимі drone виконує поставлені перед ними завдання без додаткової участі оператора. Прикладами такого режиму є: керування двигуном drone; автоматичне керування drone, яке має бути активоване або деактивоване оператором drone. Четвертий рівень автономії – це повністю автономний drone. Такий drone приймає команди, що вводяться оператором, і переводить їх у конкретні завдання без подальшої взаємодії з оператором. Але у разі виникнення нештатної ситуації оператор може втрутитися в процес виконання drone поставленого завдання.

При визначенні до якого класу належить drone, одними з базових показників є розмір та вага drone.

Сучасні технології дозволили drones знаходитися в дуже широкому діапазоні розмірів і являти собою або невелику комаху з вагою в кілька грам – сучасний nano-drone (рис. 2), або бути порівнянним з проектом повноцінного комерційного літака з вагою сотні кілограм.

Рис. 2. Приклад nano-drone

Виходячи з цього, при визначенні класу drone додали такі параметри, як встановлений на drone двигун і джерело енергії, що використовується цим двигуном. Ці параметри впливають на:

- величину корисного навантаження, яке можна встановити на drone;
- радіус дії drone.

В даний час виділяють чотири базові джерела енергії, які використовуються двигунами drone:

- стандартне авіаційне паливо;
- багатозарядні акумуляторні батареї;
- базові паливні елементи;
- високотехнологічні сонячні панелі.

Авіаційне паливо використовується в drone, який можна порівняти з повноцінними комерційними літаками. Приклад такого drone є drone Predator (рис. 3).

Рис. 3. Приклад drone Predator

Багатозарядні акумуляторні батареї в основному використовуються в nano-drones. Такі nano-drones мають малий радіус дії і в основному призначені для проведення розвідувально-рятувальних операцій, що робить їх ефективними помічниками. Базові паливні елементи, які використовуються двигунами drones - це пристрої, що виконують перетворення фізичної енергії з одного стану в інший. Таке перетворення відповідає вимогам green energy. Однією з переваг використання базових паливних елементів є великий радіус дії drones, а відповідно і час польоту. Наприклад, drone Stalker, в якому використовується базовий паливний елемент, може виконувати поставлені перед ним завдання протягом 10 годин, замість 1.5 години при використанні nano-drone на акумуляторах. Використання високотехнологічних сонячних панелей у сучасних drones в даний час зустрічається рідко. Проте, останні досягнення у галузі високих технологій та вимоги green energy, дозволили світовим компаніям Google та Facebook вкласти свої гроші в дослідні проекти з експериментального виробництва drones з використанням високотехнологічних сонячних панелей. На сьогодні завдання, які вирішуватимуть такі drones, це надання бездротового підключення користувачів до високошвидкісної мережі Internet у важкодоступних місцях [3]. Перевагами таких drones є можливість тривалий час перебувати в повітрі і не забруднювати physical environment. При виконанні аналізу поставлених завдань перед drones, до складу якої входять drones, порівняні з проектом повноцінного комерційного літака з вагою в сотні кілограм, була виявлена можливість виникнення нештатних ситуацій, обумовлених відмовами функціональних підсистем таких drones. Традиційно стійкість технічних систем до виникнення нештатних ситуацій обумовлюється властивістю надійності та передбачає дублювання окремих бортових підсистем [3]. Однак, для підсистем drones даний підхід не застосовується через обмеження за масою, габаритами та обчислювальними можливостями. В даний час розробниками у підсистемах drone використовується функціональна надмірність. Вона дозволяє відбивати наслідки позаштатних ситуацій з допомогою реконфігурації. Тому при проектуванні drones, розробники застосовують підсистеми з програмованою логікою та структурою, що перебудовується. Реконфігурація дозволяє забезпечити функціонування drones у разі виникнення нештатних ситуацій. Реконфігурація дає можливість забезпечити живучість drone та виконати поставлені перед ним завдання у разі виникнення нештатних ситуацій [3].

Аналіз останніх досліджень та публікацій. Колективом дослідників – Barabash O., Tverdenko H., Sobchuk V., Musienko A., Lukova-Chuiko N. були опубліковані результати, які показують, що при появі зовнішніх або внутрішніх деструктивних впливів на «Swarm-bot» systems Найбільш ефективним рішенням для досягнення поставленого завдання (підвищення живучості) є застосування одночасно групи інтелектуальних мобільних «s-bots», що входять до складу одної «Swarm-bot» system [4]. Провідними вченими України – Dodonov, A.G., Gorbachyk, O.S., Kuznietsova, M.G., опубліковано роботи, з яких видно, що при використанні інтелектуальних мобільних «s-bots», оснащених автономною системою пересування та навігації та здатних до виконання певних функцій, виникають складні завдання, пов'язані насамперед із проблемою управління такими засобами та організацією колективної їх взаємодії для найбільш ефективного досягнення поставленого завдання [5].

Мета статті. Провести дослідження підходів до розрахунку кількісної оцінки живучості drones на етапі проектування.

Виклад основного матеріалу

Огляд розрахунку кількісної оцінки живучості drones. В даний час термін живучість включає в себе два базові параметри: надійність та безпека. Живучість як окремих підсистем drone, так і drone загалом характеризується базовими властивостями: невразливістю; стійкістю; реінжинірингом. Тому можна з упевненістю сказати, що живучість відображає стійкість до дистуктивних впливів як окремих підсистем drone, так і drone загалом [6]. Така живучість закладена в алгоритмічну частину підсистем drone і дозволяє у разі виникнення нештатних ситуацій змінювати послідовність роботи підсистем drone [7]. При вирішенні задачі аналізу та синтезу елементів підсистем drone, оперують кількісною оцінкою живучості [8]. Це необхідно для визначення, з якою ймовірністю відмови окремих елементів будь-якої підсистеми drone призведуть до пошкодження всього drone [9]. В даний час серед розробників drones не існує єдиної думки для визначення кількісної оцінки живучості drones [10]. Багато залежить від архітектури підсистем drone. Їх взаємодії між собою та взаємодії їх елементів, між собою [11].

Перший підхід до розрахунку кількісної оцінки живучості drones. Розглядаючи кількісну оцінку живучості drone в цілому, враховують особливості архітектури підсистем drone та їх взаємодії між собою. Досліджується зміни станів підсистем drone у часі: перший стан – початковий стан підсистем drone – S_S та початковий момент часу – t_S – S_S (t_S); другий стан – кінцевий стан підсистем drone –S_D та кінцевий момент часу – t_D – S_D (t_D).

При виникненні диструктивної дії I_N на підсистеми drone, може призвести до того, що запланований кінцевий стан S_D не буде досягнуто. При цьому підсистеми drone опиняться у поточному стані $S_T(t_T)$. Така ситуація може виникнути, коли почнуть послідовно відмовляти підсистеми drone. У такому випадку живучість drone характеризується

максимальною кількістю елементів, які відмовили, що входять до складу підсистем drone – E_{NW} , після яких зберігається працездатність drone:

$$G = \max_{\prod S}(E_{NW}), \quad (1)$$

де E_{NW} – кількість елементів, які відмовили та входять до складу підсистем drone.

Другий підхід до розрахунку кількісної оцінки живучості drones. При розрахунку кількісної оцінки живучості drone враховується функціональний підхід, тому що підсистеми drone характеризуються: переліком виконуваних завдань; цільовим функціоналом; безліччю елементів. Елементи підсистем drone можуть знаходитись у трьох базових станах: в робочому стані; не робочому стані; частково робочому стані. Тоді здатність drone до функціонування, після деструктивного впливу, залежить від вимог, що висуваються до виконуваних завдань. У загальному переліку розробники drone виділяють такі вимоги: вимога перша – необхідно виконати всі завдання, навіть якщо це призведе до втрати якості виконання; вимога друга – виконати пріоритетні завдання; вимога третя – виконати одне з базових завдань.

Вимога перша. Визначається розрахункове значення ефективності функціонування drone – *W*, яке розраховується залежно від специфіки роботи drone.

Вимога друга. Визначається кількість пріоритетних завдань $\overline{task_{priority}}$, виконуваних drone в умовах виникнення відмов елементів підсистем drone по відношенню до базової кількості завдань – *task* :

$$G = \overline{task_{priority}} \big/ task, \quad (2)$$

Вимога третя. Виконати одну з базових задач, поставлених перед drone, при виникненні відмови хоча б одного елемента, що входить до складу підсистем drone:

$$G = P_{subsystem} \big/ P_{element}, \quad (3)$$

де $P_{subsystem}$ – ймовірність відмови підсистеми drone; $P_{element}$ – ймовірність відмови елемента, що входить до складу підсистеми drone.

Третій підхід до розрахунку кількісної оцінки живучості drones. Розробниками drones оцінюється мінімальна кількість зв'язків між елементами, які входять до складу підсистем drone і які необхідно розірвати, щоб drone припинив виконувати поставлені перед ним завдання:

$$G = \min_{\prod S}(L_E : W_T = 0), \quad (4)$$

де L_E – зв'язок між елементами, що входять до складу підсистем drone; W_T – ефективність функціонування drone в даний момент часу – t_T.

Четвертий підхід до розрахунку кількісної оцінки живучості drones. Розробниками drones розглядається логіко-імовірнісний підхід до оцінки живучості. Безпосередньо виділяється бінарна логіка функціонування підсистем drone та її елементів. Сам drone описується за допомогою математичного апарату – статичної моделі, а функціональні зв'язки між елементами, що входять до складу підсистем drone, представляються за допомогою алгебри логіки. Оцінку живучості розробники drones виконують за допомогою імітаційного моделювання, задаючи різні комбінації відмов елементів, що входять до складу підсистем drone. Після цього проводяться обчислення логічної функції, за результатами яких визначається здатність drone виконувати поставлені перед ним завдання. Далі розробники drones визначають безліч, яка складається з елементів, що відмовили – E_{NW} . Така безліч E_{NW} впливає на якісне виконання поставлених перед drone завдань, тому ці відмови називаються – критичними. Тоді показником живучості можна вважати, як ймовірність збереження drone здатності виконувати поставлені перед ним завдання при *k*-кратному виникненні відмов і розрахована як відношення E_{NW} кількості відмов елемента, що призвело до порушення роботи drone, до загальної кількості відмов $E_{\sum NW}$ цього елемента:

$$G = \sum_{i=1}^{N} \frac{k_i (1 - E_{NW})}{E_{\sum NW}}, \quad (5)$$

де k_i – ваговий коефіцієнт; E_{NW} – кількість елементів, які відмовили та входять до складу підсистем drone; $E_{\sum NW}$ – загальна кількість відмов цього елемента.

Сучасний підхід до розрахунку кількісної оцінки живучості drones. В даний час розробниками складних технічних систем запропоновано більш простий підхід до обчислення кількісної оцінки живучості як підсистем drone, так і самого drone. У такій послідовності дій необхідно знати, чи підсистеми drone зможуть виконувати базовий набір функцій при виникненні деструктивного впливу. Для цього розробники drone на етапі проектування закладають у проект drone функціональну надмірність, для того щоб при виникненні деструктивного впливу drone не вийшов з ладу. Тоді можна говорити про те, що drone виконає в повному обсязі поставлені перед ним завдання, тобто про не повний вихід підсистем drone з ладу, а відтак і кількість працездатних елементів – E_W , що входять до їх складу. Тому, при обчисленні кількісного показника живучості drone, в даний час розробники використовують відношення ефективності функціонування drone в поточний момент часу – $W_T(t_T)$, до ефективності функціонування drone у початковий момент часу – $W_0(t_0)$:

$$G(t) = W_T(t_T) / W_0(t_0), \quad (6)$$

де W_T – ефективність функціонування drone в поточний момент часу – (t_T); W_0 – ефективність функціонування drone у початковий момент часу – (t_0). Тоді для математичної моделі, яка дає можливість обчис-

лення кількісного показника живучості drone, формується такий перелік припущень та обмежень:

- кількісний показник ефективності функціонування елементів, що входять до складу підсистем drone, визначається на основі базових характеристик елемента – E_W ;

- математична модель, яка дає можливість обчислення кількісного показника живучості drone актуальна, доки не виникне відмова елементів – E_{NW} , що входять до складу підсистем drone. Потім математична модель піддається коректуванню.

Виходячи з переліченого переліку припущень та обмежень, обчислення кількісного показника живучості drone виконується на базі аналізу якісного виконання всіх поставлених перед drone завдань. При цьому визначення ефективності функціонування drone W_T у поточний момент часу (t_T) обчислюється знаходженням добутків показників ефективності виконання поставлених перед drone завдань для загальної кількості працездатних елементів E_W , що входять до складу підсистем drone.

Висновки

Підсистеми drone, та й сам drone, являють собою складну технічну систему з програмованою логікою та структурою функціонування всіх елементів, що перебудовується. Завдяки програмованій логіці, закладений в елементи алгоритм роботи підсистем drone, дозволяє якісно виконувати поставлені перед drone базові завдання. При цьому алгоритми, які існують для кількісної оцінки живучості, включають різні сутності опису підсистем drone, як складної системи. В даний час розробниками drone запропоновано виконувати кількісну оцінку живучості drone, яка базується на аналізі ефективності роботи елементів, що входять до складу підсистем drone, при виникненні деструктивної дії. Це дозволяє оцінити можливість drone виконати поставлені перед ним базові завдання і, при необхідності, визначити потребу в резервних шляхах (перебудовувана структура) взаємодії між елементами, що входять до складу підсистем drone.

СПИСОК ЛІТЕРАТУРИ

1. Koshovyi M. D., Pylypenko O. T., Ilyina I. V., Tokarev V. V. Growing tree method for optimisation of multifactorial experiments, Radio Electronics, Computer Science, Control, 2023, № 3, pp. 55–61. Doi: 10.15588/1607-3274-2023-3-6.
2. Кривуля Г.Ф., Токарєв В.В., Ільїна І.В., Кравець В.Є. Взаємодія між «s-bots» однієї «Swarm-bot» system у фізичному неорганізованому середовищі, Системи управління, навігації та зв'язку, 2023, №1(71), с.108-111. Doi: 10.26906/SUNZ.
3. Krivoulya G., Koshevoy N., Tokariev V., Ilina I., Dubinsky D. Solving the Task of Topological Formation Intelligent Mobile «S-bots» for One «Swarm-bot» System, Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems: (COLINS 2023). CEUR Workshop Proceedings, 2023. Kharkiv. Ukraine, pp. 273-282.
4. Krivoulya G., Tokariev V., Ilina I., Lebediev O., Shcherbak V. Algorithm of Iterations of Distribution of Subtasks Between «S-Bot» in One «Swarm-Bot» System, Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Systems: (COLINS 2022). CEUR Workshop Proceedings, 2022. Gliwice. Poland, pp. 1531-1541.
5. Krivoulya G., Ilina I., Tokariev V., Shcherbak V. Mathematical Model for Finding Probability of Detecting Victims of Man-Made Disasters Using Distributed Computer System with Reconfigurable Structure and Programmable Logic, PIC S&T, 2020. Kharkiv. Ukraine, pp.573 - 576.
6. Churyumov G., Tkachov V., Tokariev V., Diachenko V. Method for Ensuring Survivability of Flying Ad-Hoc Network Based on Structural and Functional Reconfiguration, ITS 2018, CEUR Workshop Proc., 2018. Kyiv, Ukraine, pp. 64-76.
7. Churyumov G., Tokarev V., Tkachov V., Partyka S. Scenario of Interaction of the Mobile Technical Objects in the Process of Transmission of Data Streams in Conditions of Impacting the Powerful Electromagnetic Field, DSMP, 2018. Lviv, pp.183-186.
8. Ткачов В.М., Токарев В.В., Радченко В.О., Лебедєв В.О. Проблема передачі даних типу BIG DATA у мобільній системі «Мультикоптер-сенсорна мережа», Системи управління, навігації та зв'язку, 2017, №2(42), с.154-157.
9. T. Gao, X. Bai. Bayesian. Optimization-based Three-dimensional, Time-varying Environment Monitoring using an UAV, IEEE Journal of Intelligent & Robotic Systems, 2022, vol.105, no.4, pp.219 - 235. Doi:10.1007/s10846-022-01709-x.
10. L. Zhao, Y. Song, C. Zhang. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction, IEEE Transactions on Intelligent Transportation Systems, 2020, vol.21, no.9, pp.3848 - 3858. Doi:10.1109/TITS.2019.2935152.
11. E. Seraj, A. Silva, M. Gombolay. Multi-UAV planning for cooperative wildfire coverage and tracking with quality-of-service guarantees, Autonomous Agents and Multi-Agent Systems, article number.39, 2022, Springer. Doi:10.1007/s10458-022-09566-6.

Received (Надійшла) 05.11.2023
Accepted for publication (Прийнята до друку) 24.01.2024

Analysis of existing approaches to calculating quantitative assessment of drones survivability

O. Lebediev, O. Bondar, E. Samoilenko, V. Cherevko

Abstract. Today, survivability reflects the resistance of both individual drone subsystems and the drone as a whole to destructive influences. Such survivability is built into the algorithmic part of the drone subsystems and allows, in the event of emergency situations, to change the sequence of operation of the drone subsystems. When solving the problem of analyzing and synthesizing elements of drone subsystems, they operate with a quantitative assessment of survivability. This is necessary to determine with what probability failures of individual elements of any drone subsystem will lead to damage to the entire drone. Currently, there is no consensus among drone developers to quantify the survivability of drones. A lot depends on the architecture of the drone subsystems. Their interactions with each other and the interactions of their elements with each other. This article is devoted to the study of approaches to calculating a quantitative assessment of the survivability of drones at the design stage.

Keywords: "Swarm-bot" - system, "s-bots", survivability, safety, reliability, invulnerability, resilience, reengineering, communication, nano-drone.

УДК 004.896 **doi: 10.26906/SUNZ.2024.1.122**

С. Ю. Леонов, Д. А. Тиртишний

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

ДОСЛІДЖЕННЯ ПРОДУКТИВНОСТІ СЕРВЕРНОЇ ЧАСТИНИ КОМП'ЮТЕРНОЇ СИСТЕМИ НА ОСНОВІ РОЗРОБЛЕНОГО ФРЕЙМВОРКУ

Анотація. У даній статті були вивчені методи та використання оригінально розробленого фреймворку для тестування продуктивності серверної частини комп'ютеної системи. Було проаналізовано завдання та виклики, які стоять перед розробниками при оптимізації продуктивності серверів. Проведено моделювання роботи сервера та створено фреймворк, що включає інструменти для оптимізації навантаження, моніторингу, аналізу даних, керування тестовим процесом та розподілу ресурсів. Експерименти проведено на реальних веб-додатках, що використовувалися як тестове середовище. Система прийняття рішень на основі запропонованого фреймворку призводить до оптимізації продуктивності серверної частини веб-додатків, що значно поліпшує їхню працездатність і надійність. В ході досліджень були визначені та підтверджені ознаки подальшого покращення продуктивності веб-додатків.

Ключові слова: тестування продуктивності; комп'ютерна система; веб-додатки; фреймворк; оптимізація; розподіл ресурсів; система прийняття рішень.

Вступ

Тестування продуктивності веб-сайту — це засіб забезпечення якості, який передбачає використання автоматизованого тестування, що імітує роботу певної кількості бізнес користувачів з їх транзакціями [1, 2]. Цей вид тестування є обов'язковим у процесі покращення якості кінцевого продукту [3].

Причина полягає в тому, що будь-якого роду перебої або проблеми, пов'язані з поганою продуктивністю-можуть стати причиною відмови клієнтів від використання конкретного програмного забезпечення [4].

Розглянемо ситуацію: чорна п'ятниця, день скидок, інтернет-магазин розраховує на велику кількість продажів у цей день. І, раптом, сервер відмовляє, перестає працювати, усі люди виходять з сайту та починають робити покупки в магазинах конкурентах. Скільки грошей втратила би компанія AliExpress, якби їх сервера відмовили у чорну п'ятницю. Тобто, стабільність роботи є чи не найбільш важливішим атрибутом Web-сайту, а тестування продуктивності створене саме для забезпечення цієї стабільності. Таке тестування дає можливість переконатися, що випробуване програмне забезпечення чи аплікація добре працюють при критичних умовах [5] та допомагає визначити, наскільки швидко деякі конкретні частини її системи реагують у найгірших умовах [6].

Мета статті. Виконати дослідження та розробити фреймворк для тестування продуктивності серверної частини веб-додатків.

Основна частина

В залежності від характеристик системи, які підлягають перевірці, виділяють декілька типів тестування продуктивності [7, 8]:

- Performance Testing (Тест продуктивності) — це будь-який тест, який перевіряє стабільність, продуктивність, масштабованість та / або пропускну спроможність веб-сайту.
- Capacity Test, Volume Testing (Тест на місткість) — допомагає визначити, скільки користувачів / об'єм їх даних може обробляти веб-сайт або додаток, перш ніж продуктивність або стабільність стають неприйнятними.
- Load Test (Тест на завантаження) — це тестування реакції системи на зміну навантаження (в межах допустимого). Різниця із Performance Testing (Тест продуктивності) в тому, що навантаження може вимірюватися навмисне не на піку.
- Stress Test (Стрес-тест) — як показує його назва, подібне тестування приведе до того, що Ваша програма виконуватиметься за ненормальних умов. Це дозволить дізнатись, її здатність до регенерації, після завершення дії стресу. Стресове тестування також показує, які компоненти зникнуть на крайньому рівні.
- Stability Test(Тест стабільності на витривалість) — це тривалий тест, який використовується для оцінки продуктивності та / або стабільності програми у часі. Корисний, тому що при проведенні цього виду тестування здійснюється спостереження за споживанням пам'яті для виявлення потенційних витоків. Крім того, таке тестування виявляє деградацію продуктивності, котра виражається у зниженні швидкості обробки інформації та / або збільшенні часу відповіді аплікації після тривалої роботи, порівняно з початком тесту [9].
- Smoke Test (Димовий тест) — це короткі цикли тестів, які проводяться під дуже низьким навантаженням. Цей вид продуктивного тестування підкреслює, що програма працює як очікується. Цей термін походить від апаратного тестування, де, якщо дим генерується (буквально), це означає, що тест не вдався, і більше тестування не потрібне.

Для реалізації фреймворку тестування продуктивності серверної частини потрібно спочатку дослідити різні інструменти для тестування продуктивності, вивчити їх взаємодію та розробити архітектурну діаграму (рис. 1). Для того щоб забезпечити процес неперервної підтримки тестів та зрозумілий інтерфейс для їх запуску, була використана зв'язка Jenkins + Git. У [10, 11] описані Jenkins та Git. У фреймворку тестування продуктивності Git використовується для того щоб зберігати тестові файли та файл з параметрами запуску тестів.

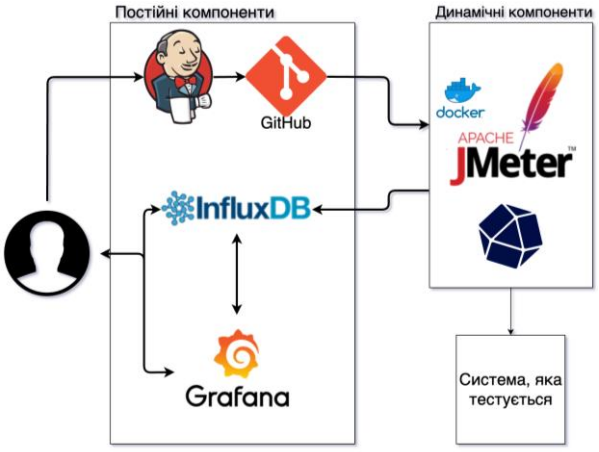


Рис. 1. Високорівнева архітектура фреймворку тестування продуктивності

Jenkins у свою чергу, надає можливість запуску тестів через зрозумілий графічний інтерфейс та з можливістю вибору параметрів запуску: кількість віртуальних користувачів, тип тесту, тестовий сценарій, тощо (рис. 2).

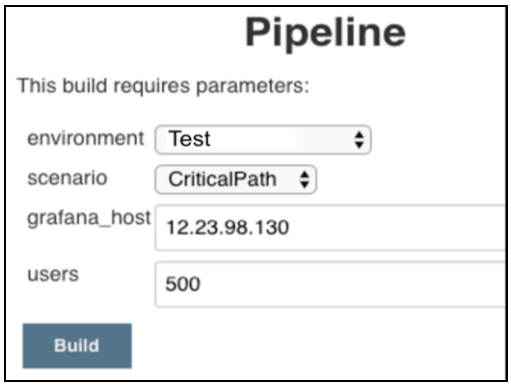


Рис. 2. Параметри запуска тесту в Jenkins

Детальний опис параметрів:

- environment – вибір системи, над якою буде проводитись тестування;
- scenario – вибір тестового сценарію, який буде протестований;
- grafana_host – адреса бази даних, яка зберігає результати та робить їх візуальне відображення (у вигляді графіків та таблиць);
- users – кількість віртуальних користувачів, які будуть робити запити до серверу одночасно.

Apache JMeter — інструмент для проведення навантажувального тестування, що розробляється Apache Software Foundation, підпроекту Jakarta. Хоча спочатку JMeter розроблявся як засіб тестування веб-застосунків, натепер він здатний проводити навантажувальні тести для JDBC-з'єднань, FTP, LDAP, SOAP, JMS, POP3, IMAP, HTTP і TCP.

Цікава можливість – створення великої кількості запитів за допомогою декількох комп'ютерів при управлінні цим процесом з одного з них. Архітектура підтримує плагіни сторонніх розробників і дозволяє доповнювати інструмент новими функціями.

Docker — інструментарій для управління ізольованими Linux-контейнерами. Docker доповнює інструментарій LXC більш високорівневим API, що дозволяє керувати контейнерами на рівні ізоляції окремих процесів. Зокрема, Docker дозволяє не переймаючись вмістом контейнера запускати довільні процеси в режимі ізоляції і потім переносити і клонувати сформовані для даних процесів контейнери на інші сервери, беручи на себе всю роботу зі створення, обслуговування і підтримки контейнерів.

Сирцевий код Docker написаний мовою Go і поширюється під ліцензією Apache 2.0. Інструментарій базується на застосуванні вбудованих в ядро Linux штатних механізмів ізоляції на основі просторів імен (namespaces) і груп управління (cgroups). Для створення контейнерів використовуються скрипти lxc. Для формування контейнера досить завантажити базовий образ оточення (команда docker pull base), після чого можна запускати в ізольованих оточеннях довільні програми (наприклад, для запуску bash можна виконати docker run -i -t base/bin/bash).

Docker дозволяє нам динамічно створювати генератор навантаження, у вигляді jmeter. Це має ряд переваг: у разі якщо виникли проблеми с генератором навантаження ми можемо легко перезапустити його за допомогою створення нового docker контейнеру. Також це дозволяє створювати генератори навантаження із потрібною нам конфігурацією відразу, тобто нам не потрібно буде кожен раз налаштовувати усю систему, вона буде відразу налаштована.

Telegraf – це додаток, який записує усі параметри моніторингу системи до бази даних, які потім можна буде використовувати для створення графіків у Grafana. Що дозволить нам бачити, яка частина комп'ютера дала збій під час тесту: процесор, оперативна пам'ять, диск чи з'єднання з інтернетом. Тобто Telegraf дозволяє робити моніторинг системи-сервера, яка тестується, у реальному часі.

InfluxDB – це база даних на основі часових рядів у якій зберігаються результати наших тестів. Усі дані зберігаються у вигляді пари ключ-значення, де ключем є поточний час. Також у цю базу можливо записувати спеціальні теги, які допоможуть фільтрувати дані та групувати їх на Grafana.

Grafana – це додаток для візуалізації даних, які зберігаються в InfluxDB. Найголовніший візуальний об'єкт у Grafana – це dashboard. Dashboard-це набір рядків, в кожному з яких є одна чи декілька панелей. Панелі бувають різні: таблиці, цифрові панелі, повідомлення, кругові діаграми, теплові карти, тощо. Також є можливість дивитись результати за вибраний період часу, що також є дуже важливою можливістю.

Перед тим як почати розробку, потрібно зрозуміти як компоненти будуть взаємодіяти між собою. Почати потрібно з установки Docker, далі будуємо "Docker image"(образ докер) для Grafana та Influx. Образ Docker містить операційну систему, застосунок і всі його залежності. Образи в Docker складаються з шарів. Якщо треба отримати образ з веб-сервером, то беремо за основу образ з дистрибутивом операційної системи, додаємо залежність - веб-сервер, і записуємо це як новий образ, який матиме два шари – один з ОС, наступний з веб-сервером. Образами можна обмінюватись через DockerHub.

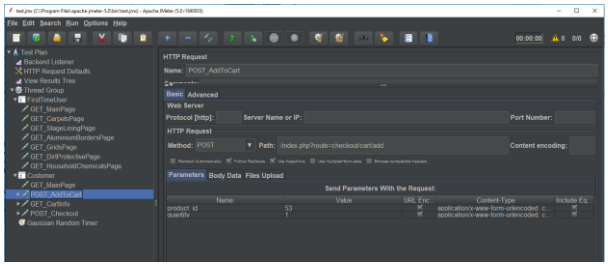
Саме з DockerHub можна завантажити офіційні образи InfluxDB, Grafana та Jmeter.

Після того, як образи були завантажені, їх потрібно встановити, або запустити. Запущені образи докерів називаються Docker Container. Запустимо контейнери бази даних InfluxDB та візуального додатку Grafana. Як тільки вони будуть запущені – ви зможете перейти по адресі машини на якій була запущена Grafana та продовжити настройку там. Приклад URL: https://localhost/grafana . Ці контейнери ми запускаємо відразу, бо вони є постійними компонентами. Доступ до них ми повинні мати постійно, аби завжди мати змогу подивитися результати попередніх тестів. Образ Jmeter буде запускатися динамічно, одразу після кліку на кнопку “build” у Jenkins.

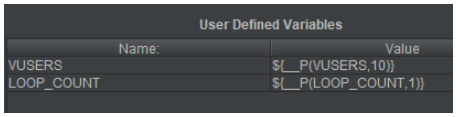
Наступним етапом буде установка Jmeter на свій комп’ютер для скриптування запитів до серверу на основі тестового сценарію. Jmeter складається з декількох основних компонентів: Thread Group, HTTP Requests, Listeners та таймери [20].

- Thread Group: налаштування усього сценарію, в яких вибираються кількість віртуальних користувачів, кількість запусків тестового сценарію;
- HTTP Requests: це саме наші запити до сервера, в них ми вказуємо URL, за яким ми відправляємо запит, тип запиту, та параметри;
- Listeners: слухачі, які запам’ятовують та агрегують результати тесту, в нашому випадку використовуємо спеціальний Backend Listener який буде писати результати в базу даних InfluxDB;
- Таймери дозволяють встановлювати паузи між запитами, щоб емулювати більш реальний сценарій використання сайту, бо навряд чи ви змогли би передивитися 5 сторінок сайту за 3 секунди.

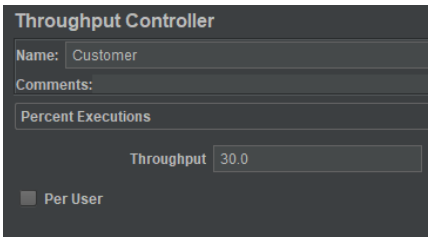
Інтерфейс Jmeter, та усі позначені вище його головні компоненти для тестового сценарію наведені на рис. 3, а.



а



б



в

Рис. 3. Тестовий сценарій у Jmeter

На рис. 3, б можна побачити 2 змінні VUSERS та LOOP_COUNT. Вони позначають кількість одночасних віртуальних користувачів та кількість запусків сценарію. Ці 2 змінні мають значення у виді змінної та значення за замовчанням. Це зроблено для того, щоб можна було передавати параметри при запуску тесту через файл з параметрами, або через Jenkins. Але якщо ніяких параметрів не передавати, то буде взято значення за замовчанням.

На рисунку 3, в можна побачити яким чином було реалізоване розділення кількості користувачів на конкретні запити до сервера. Був використаний пропускний контролер, котрий пропускав лише 30% від загальної кількості віртуальних користувачів.

Після того як тести були успішно розроблені у Jmeter, час додати їх на GIT та з'єднати Git репозиторій із Jenkins проектом. Після того як усе це зроблено залишається лише додати Pipeline Script до Jenkins проекту, який буде доставати тести з репозиторія, після цього створювати файл з параметрами, записувати туди параметри для тестів, вказані користувачем при запуску, далі він створює контейнер з Jmeter, в який додається тест та файл параметрів, після того тести проходять усередині контейнеру, контейнер видаляється, та результати відображаються у Grafana (рис. 4). Скрипт має бути написаний на мові Groovy.

```
stage('Execute tests') {

    sh "docker run --rm --volume
      `pwd`/Tests:/mnt/jmeter
      jmeter:latest -n -t
      /mnt/jmeter/${scenario}/test.jmx
      -q /mnt/jmeter/${scenario}/parameters.txt"
```

Рис. 4. Команда запуску контейнера Jmeter

Після цього залишається лише запустити тести та аналізувати результати.

Почнемо з тесту під навантаженням в 100 одночасних віртуальних користувачів (рис. 5)

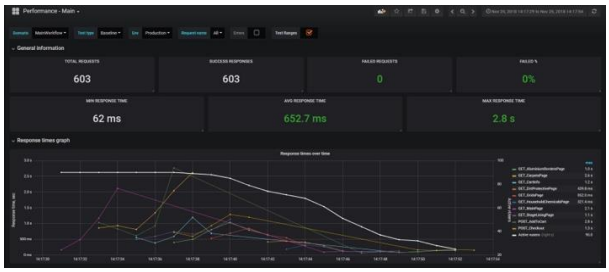


Рис. 5. Тестовий сценарій
100 віртуальними користувачами

На рисунку показано, що зі 100 одночасними користувачами система справляється дуже добре, максимальний час відповіді сервера клієнту склав 2.8 секунди, а середній лише 652 мс. На графіку, спочатку різко почала заходити велика кількість користувачів на сайт, тому час відповіді був доволі високий, але поступово користувачі закінчували свою роботу й показники покращувались. Верхня

біла лінія відображає кількість віртуальних користувачів у секунду часу. На піку, 90 користувачів в одну секунду запитували сервер. Для більш детального розбору кожного запиту потрібна детальна таблиця (рис. 6).

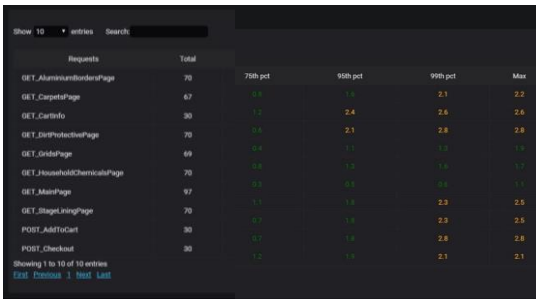


Рис. 6. Таблиця для сценарію зі 100 користувачами

Далі було прийняте рішення знайти кількість одночасних користувачів ближчу до максимуму, тест на місткість. Проведемо тестування з 200 одночасними користувачами (рис.7).

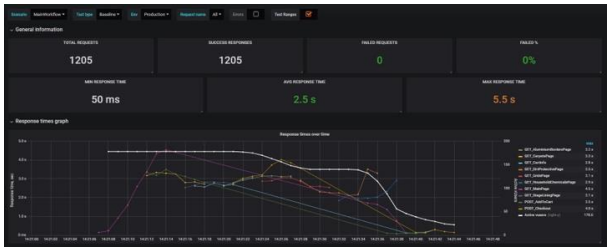


Рис. 7. Тест на місткість,
200 одночасних віртуальних користувачів

Середня швидкість відповіді сервера клієнту складає 2.5 секунди, що є нормою, але максимальний час вже виходить за норму, тобто можна вважати, що 200 користувачів є останньою межею перед початком нестабільної та повільної роботи сервера.

Наступним видом тестування було прийняте рішення зробити стресове тестування. Такий вид тестування допоможе нам побачити помилки, які викидає сервер, та навантаження на нього.

Проведемо тестування з 500 одночасними користувачами. (рис. 8)

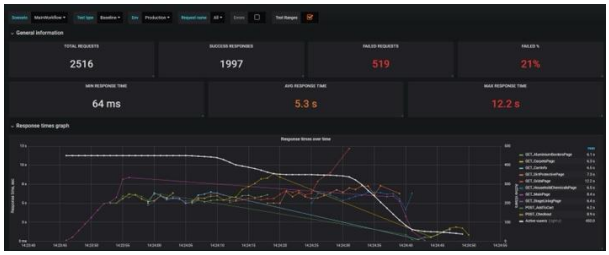


Рис. 8. Стрес тест, 500 віртуальних користувачів

При такому стресовому навантажені система вже стає нестабільною, та 21% запитів відбиваються з помилками. При цьому середній час відповіді сервера клієнту складає 5.3 секунди, а максимальний аж 12.2 секунди. Також можемо бачити, що найбільша кількість віртуальних користувачів у секунду, склала 450. Для того щоб розібратись з помилками, була розроблена відповідна панель з детальним описом помилок (рис. 9).

Рис. 9. Помилки при стрес тесті

Помилки були «Gateway time-out», що означає те, що клієнт не дочекався відповіді від сервера та вийшов з сесії. Це все стається у наслідок неймовірного навантаження на сервер. Після цього потрібно розібратись, а в чому все ж таки була проблема, чому сервер почав працювати повільно та нестабільно. Саме для цього нам і знадобиться інструмент моніторингу системи в реальному часі – Telegraf (рис. 10).

На рисунку вище ми бачимо, що як тільки на нашому сервері з'являється 500 одночасних користувачів, наш процесор завантажується на всі 100%, та тримає такий показник аж до кінця тесту і лише після цього, зменшує свої показники.

Рис. 10. Використання CPU сервером

Обговорення результатів

В ході дослідження були виконані дослідження з різними видами тестування продуктивності серверної частини веб-додатків. Після того були досліджені різні інструменти для розробки фреймворку для тестування продуктивності.

Після розробки фреймворку, був розроблений тест для прикладу, та завдяки цьому тестовому сценарію був протестований інтернет-магазин.

Тести продуктивності перевірили максимальну місткість сайту, його стабільність, та швидкість. Сайт може видавати відповідь на будь-який запит клієнта швидше ніж за 2.5 секунди при 100 одночасних користувачів на сайті. Також виявилось, що сайт може функціонувати навіть при 500 одночасних користувачів, але вже з деякими помилками, та затримками. Стабільна безпомилкова робота продовжується десь до 200 користувачів при нетривалому навантаженні на сервер.

При тривалому навантаженні від 100 одночасних юзерів, інтернет-магазин зберігає добрі показники швидкості, якості та стабільності. При стресових навантаженнях уся продуктивність впирається в процесор, тому порадою для власника інтернет-магазину буде зміна процесора на більш потужний, якщо він звичайно розраховує на такі великі навантаження на сервер.

Висновки

У результаті проведеного дослідження тестування продуктивності серверної частини веб-додатків та розробки фреймворку потрібно зазначити, що впровадження цього фреймворку дозволяє підвищити ефективність роботи та стабільність серверних веб-додатків.

Були детально досліджені особливості тестування продуктивності серверних веб-додатків, обговорені можливі проблеми, з якими можуть зіткнутися розробники, створені рекомендації щодо їхнього усунення.

Розроблений фреймворк включає інструменти для створення навантаження, моніторингу, аналізу даних, керування тестовим процесом, а також модулі для вдосконалення продуктивності. Він є зручним та гнучким інструментом для ефективного аналізу та вдосконалення продуктивності серверних веб-додатків.

Експерименти, проведені в рамках дослідження, підтвердили, що використання розробленого фреймворку призводить до значного покращення продуктивності серверних веб-додатків.

Отже, результати цього дослідження можуть бути використані розробниками веб-додатків для підвищення продуктивності та надійності своїх продуктів.

Список літератури

1. Леонов С. Ю., Тиртишний Д. А. Ключові аспекти тестування продуктивності. *Проблеми інформатики та моделювання* (ПІМ-2023): матеріали міжнародної науково-технічної конференції, Харків: НТУ «ХПІ», 2023. С. 69. URL: https://repository.kpi.kharkov.ua/server/api/core/bitstreams/ee691cd3-44dd-4848-8631-5df0882a6c40/content
2. Still A. Web performance warrior. The Business of Speed. Published by O'Reilly Media, Inc., CA 95472, 2015. 50 p. URL: https://theswissbay.ch/pdf/Books/Computer%20science/O%27Reilly/web-performance-warrior.pdf
3. Buckler C. Jump Start Web Performance. SitePoint, 2020. 159 p. URL: https://www.oreilly.com/library/view/jump-start-web/9781098122799/
4. Донич Д. Види тестування програмного забезпечення. 2021. URL: https://lemon.school/blog/vydy-testuvannya-programnogo-zabezpechennya
5. Devaraj K. Software Testing Models. What it is, Types & How They Work? 2023. URL: https://testsigma.com/blog/software-testing-models/
6. Dustin E., Rashka J., Paul J. Automated Software Testing: Introduction, Management, and Performance. Addison-Wesley Prof., 1999. 608 p. URL: https://books.google.com.ua/books/about/Automated_Software_Testing.html?id=kl2H0G6EFf0C&redir_esc=y
7. Hoda R., Salleh N., Grundy J. The Rise and Evolution of Agile Software Devleopment. *IEEE Software*. 2018. Vol. 35, no. 5. P. 58-63. URL: https://ieeexplore.ieee.org/document/8409911
8. Dybå T., Dingsøyr T. Agile Project Management: From Self-Managing Teams to Large-ScaleDevelopment. *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*. Florence, 2015. https://ieeexplore.ieee.org/document/7203121
9. Whiting E., Datta S. Performance Testing and Agile Software Development: A Systematic Review. 2021. 36 p. URL: https://www.researchgate.net/publication/351410867_Performance_Testing_and_Agile_Software_Development_A_Systematic_Review
10. Loadster. URL: https://loadster.app/guides/front-end-vs-back-end-performance/#the-8020-rule-of-web-performance
11. Носков В. І., Тиртишний Д. А. Тестування продуктивності web-сайту для зростання якості продукту. *Інформатика, управління та штучний інтелект* : матеріали 5-ї міжнар. наук.-техн. конф. студентів, магістрів та аспірантів, 20-22 листопада 2018 р. / наук. ред. В. Д. Дмитрієнко ; Нац. техн. ун-т "Харків. політехн. ін-т". Харків : НТУ "ХПІ", 2018. С. 65. URL: https://repository.kpi.kharkov.ua/server/api/core/bitstreams/6476d887-10a4-48af-9b25-75a9e2e68856/content

Received (Надійшла) 12.12.2023
Accepted for publication (Прийнята до друку) 07.02.2024

Research of server-side performance testing and framework development

Sergey Leonov, Dmytro Tyrtyshnyi

Abstract. This article examines the methodologies and application of a unique framework developed for performance testing of the server-side of web applications. The academic study delves into the tasks and challenges developers commonly face while seeking to optimize server performance. Server operation was simulated and a comprehensive framework was developed. This framework incorporates comprehensive tools to create a load, perform monitoring functions, analyze data, and manage the test process. All experiments for this study were implemented on actual web applications providing a real-world test environment. The study shows that implementing the proposed framework optimizes the performance of the server-side of web applications. This significant improvement enhances operability and reliability. Our research has also narrowed down and confirmed specific indications that could lead to further improvements in web application performance.

Keywords: performance testing, server-side, web applications, framework, optimization, experiment, development.

УДК 004.7 **doi: 10.26906/SUNZ.2024.1.127**

О. С. Ляшенко, І. А. Великодний, В. Г. Знайдюк, О. Д. Журило

Харківський національний університет радіоелектроніки, Харків, Україна

МОДЕЛЬ ТА МЕТОДИ ВИЯВЛЕННЯ ШИРОКОМАСШТАБНОЇ АТАКИ В СЕРЕДОВИЩІ IoT

Анотація. Головною концепцією і предметом дослідження є виявлення різного типу обширних атак в інфраструктурі IoT, огляд представленої моделі, методів та існуючих передових систем виявлення вторгнень. **Метою** даної роботи є запропонування системи виявлення вторгнень в режимі реального часу, яка буде навчена на наборі з великим обсягом даних, за допомогою нейронної мережі з використанням ансамблевого **методу** машинного навчання. **Предметом дослідження** є огляд існуючих методів та моделей виявлення широкомасштабної атаки та запропонування власного рішення системи виявлення вторгнень, яка буде базуватися на методі виявлення аномалій та нейронної мережі. **Висновок.** Побудована система виявлення вторгнень, яка аналізує інтернет трафік, вилучає ознаки з пакету, обробляє їх та передбачує різні види атак, а також характеризує їх за типом. Загрозу безпеці можна вважати основною критичною проблемою для пристроїв IoT, тому використання таких систем зменшує ризики втрати даних.

Ключові слова: набір даних, нейронна мережа, машинне навчання, мережевий трафік, IDS, навчання, передбачення, виявлення аномалій, атака, Інтернет речей, система виявлення вторгнень.

Вступ

Інтернет речей – концепція мережі, яка об'єднує фізичні пристрої з вбудованими датчиками, а також програмним забезпеченням, що забезпечує ефективну та спрощену взаємодію між фізичним світом і комп'ютерними системами, за допомогою, найчастіше, стандартних протоколів зв'язку. Протягом останніх років він стрімко зростав та продовжує зростати у різних галузях. Пристрої IoT функціонують у сферах освіти, охорони здоров'я, сільському господарстві, транспортних системах та промисловості. Кількість підключених пристроїв по всьому світу стрімко росте. Системи включають в себе масу датчиків, які дозволяють збирати дані в реальному часі. Отримані дані, це свого роду фундамент для створення інтелектуальних алгоритмів прийняття рішень. Ростуча кількість пристроїв, ціна і важливість інформації збільшує ризик кіберзагроз і викраденню інформації в корисних цілях. Виходячи з цього розробка інтелектуальних методів та систем виявлення вторгнень для пристроїв IoT стає необхідною для їх ефективного захисту. Тема безпеки інформаційного середовища стає дедалі актуальною і кібербезпека набуває життєвої важливості, з огляду на те що IoT є драйвером промислової революції та системою для збору живих даних [1]. Таким чином, система виявлення вторгнень є необхідною для виявлення і захисту мережі та пов'язаних систем від поточних і майбутніх кібератак.

Системи виявлення вторгнень

Визначимо концепцію IDS (*Intrusion Detection System, або система виявлення вторгнень*). Це програмний або апаратний засіб, який виявляє або запобігає несанаційному доступу до комп'ютерної мережі чи системи. Головна мета IDS полягає в реагуванні на небезпечні події, потенційно небезпечні, або аномалії, що можуть вказувати на вторгнення чи інші безпекові порушення. З основних завдань системи виявлення вторгнень можна виділити: *виявлення аномалій* є функцію моніторингу системи чи мережі для виявлення незвичайних патернів, подій або некоректних дій, які можуть бути ознакою вторгнення чи іншої загрози безпеці; *виявлення вторгнень* – розпізнавання несанкційного доступу, спроб атак на інформаційні системи, вірусів, троянських програм та іншого шкідливого коду; *відслідковування і реагування* – забезпечення можливості вжиття заходів до виявлених загроз, включаючи блокування доступу, відключення систем які на думку є найбільш вразливі або відправлення сповіщень адміністраторам, котрі відповідають за безпеку.

Системи виявлення вторгнень можуть використовувати різні методи, такі як сигнатурний аналіз, виявлення аномалій, використання інтелектуальних технологій, включаючи *машинне навчання* та евристичний аналіз. Ефективний захист включає в себе інтеграцію системи виявлення вторгнень з іншими методами безпеки, та поєднання цих методів, для створення комплексного захисту інформаційного стеку. Зазвичай системи виявлення вторгнень використовують два основні підходи для виявлення потенційних загроз: сигнатурний аналіз та виявлення аномалій. Розглянемо ці методи більш детально.

Сигнатурний аналіз – метод який грунтується на використанні визначених сигнатур або патернів для ідентифікації або розпізнавання конкретних відомих загроз. Сигнатури можуть представляти з себе конкретні приклади або вирази в шкідливому програмному коді, унікальні характеристики того чи іншого вірусу чи способу вторгнення, які раніше вже були визначені або вивчені [2]. Спеціалісти з безпеки аналізують атаки і розробляють сигнатури для кожного виду. Зазвичай це може бути характеристика конкретних строк коду, значень в певних полях або якийсь інший ідентифікатор, який буде унікальним для деяких типів атак. Система виявлення вторгнень застосовує ці сигнатури для пошуку вхідних даних чи активності в мережі, які відповідають зазначеним сигнатурам. Якщо є збіг, система дає сповіщення про потенційне вторгнення. Сигнатурний аналіз ефективний і має високу точність проти відомих векторів атак та відомих загроз, але не ефективний проти нових та невідомих загроз. Він потребує постійного оновлення

бази сигнатур для визначення нових загроз, більш того, зловмисники можуть уникати виявлення, шляхом зміни або шифрування свого коду. В сучасному середовищі сигнатурний аналіз залишається надійним засобом для виявлення вторгнень, але йому важко справлятися векторами атак, що постійно змінюються, які все частіше використовують нові техніки та методи, тому в сучасних *IDS* його часто доповнюють інші методи, тобто використовується комбінація різних методів для комплексного захисту, такі як виявлення аномалій, для більшої ефективності виявлення нових атак.

Виявлення аномалій – метод який базується на аналізі звичайної поведінки мережі, системи, користувачів чи інших об'єктів. Система методу будує модель так званої "норми" на основі історичних даних, фокусується на виявленні незвичайностей, відхиленню від цієї норми, яке може бути ознакою нових загроз або підозр [3]. До підходів виявлення можемо віднести: *статистичні методи,* які використовуються для аналізу величин, таких як середнє значення, середне відхилення, тощо. Відхилення від норми цих величин може вказувати на присутність аномалії; *методи машиного навчання* – створення моделей за допомогою алгоритмів машинного навчання, які можуть визначати незвичайні патерни в даних та вказувати на аномалію, наприклад, за допомогою алгоритмів кластеризації або нейронної мережі; *методи порівняння зразків* – ґрунтуються на порівнянні поточної поведінки з історичними даними, якщо виявляється відхилення від звичайної моделі, це може бути зафіксовано як аномалія.

Виявлення аномалій може проводитися на основі патернів мережевого трафіку, неправильних адрес або портів, незвичних об'ємів даних, аналіз лог файлів, що містять інформацію про дію та поведінку системи чи користувачів, надто часті або великі запити, невластиві часові рамки, тощо. Застосування методу виявлення аномалій допомагає виявляти атаки, які можуть бути невідомими(нуль-день) та непередбаченими, що робить його ефективним і корисним для захисту від нових атак та загроз [3].

Сигнатурний аналіз і виявлення аномалій часто використовують в комплексі, як частина більших систем виявлення вторгнень. Комбінація цих методів дозволяє створити більш ефективну систему виявлення вторгнень, здатну протидіяти різноманітним загрозам безпеки.

Машинне навчання в системах виявлення вторгнень

Машинне навчання відіграє важливу роль у покращені ефективності та адаптивності в системах виявлення вторгнень. Воно дозволяє системам аналізувати дані, навчатися на їх основі, та виявляти нові невідомі загрози, класифікувати події як безпечні чи підозрілі. Навчання моделі на основі історичних даних та поведінки допомагає автоматично розпізнати нові атаки чи загрози. Щоб адаптуватися до змін у поведінці системи чи користувачів системи, виявлення вторгнень можуть використовувати онлайн навчання. Це дозволить системі навчатися в реальному часі, а також підтримувати актуальність моделей. Машинне навчання ефективно працює з великими обсягами даних, що дозволяє виявляти складні патерни та взаємодії, які може бути важко виявити за допомогою традиційних методів [4–7]. Застосування машинного навчання дозволяє створювати інтелектуальні *IDS*, які можуть взаємодіяти та розпізнавати атаки на високому рівні. Використання машинного навчання в *IDS* є ключовим елементом для підвищення рівня захисту від сучасних загроз та забезпечення реактивності на нові типи атак. Розглянемо дві основні парадигми, які використовуються для розв'язання різних задач в машинному навчанні:

Supervised Learning (Навчання з вчителем) — спрямоване на розуміння зв'язку між вхідними та вихідними даними. Алгоритм, після встановлення цього зв'язку, може передбачити вихід для нових вхідних даних на основі того, що він дізнавався і зосереджується на методах класифікації та регресії. Групи класифікацій розбивають точки даних на різні класи. Цей підхід знаходить найкращій спосіб відокремити точки даних і призначити їх певним класам. Регресія відрізняється від класифікацій тим, що вона виводить число замість присвоєння точок даних класам. Класифікація фокусується на виведенні класу, тоді як регресія дає числовий вихід. Методи навчання з вчителем використовуються для виявлення відомих загроз і класифікації нових загроз за категоріями, як спам, фішинг та зловмисне програмне забезпечення [8].

Unsupervised learning (Навчання без вчителя) — набір даних містить лише вхідні дані та має справу з даними без, так званих, міток. Метою його є виявлення закономірностей або подібностей у наборі даних. Після отримання характеристик він групує дані на основі подібностей. Різниця від навчання з вчителем полягає в тому, що навчальний процес унікальний, оскільки алгоритм навчається на власному досвіді, а не на попередньо визначеному наборі вхідних даних із встановленим зв'язком. Методи навчання без вчителя використовуються для виявлення невідомих загроз і аномалій, які не належать до категорій відомих загроз [9, 10].

Оскільки кількість и складність кіберзагроз зростає, ці типи машинного навчання особливо корисні для виявлення загроз, тому що вони можуть ідентифікувати аномалії та закономірності, які можуть бути виявленні не відразу.

Концепція IDS для IoT

На сьогоднішній день концепція *IDS* застосована до IoT не є чимось новим. Було розроблено і запропоновано багато рішень і систем які використовують різні підходи та технології. Відзначимо деякі системи виявлення вторгнень для IoT:

- *Cisco IoT Threat Defense:*

Запропоноване рішення від Cisco, яке використовує аналіз трафіку, машинне навчання та інтелектуальні алгоритми для виявлення аномалій в мережі. Також вони акцентують на захисті від різноманітних атак, включаючи ті, в яких використовуються віруси та зловмисні програми.

- *Darktrace Industrial:*

Спеціалізується на застосуванні технологій штучного інтелекту для виявлення відхилень від

звичайного патерну поведінки пристроїв. Їх система враховує контекст і адаптується до змін в мережі.

- *Bastille Networks:*

Спеціалізується на безпеці радіочастотного спектру для IoT пристроїв, таких як бездротові сенсори. Вони аналізують радіохвилі для виявлення аномалій та загроз.

- *Check Point IoT Protect:*

Пропонує рішення, яке включає виявлення вторгнень для IoT пристроїв. Вони використовують технології штучного інтелекту та аналіз трафіку.

- *ARM mbed OS Sexurity:*

Виходячи з назви, надає захист на рівні ОС для IoT пристроїв через свою платформу. Вони включають заходи безпеки, такі як аутентифікація та шифрування. Важливо відзначити, що ефективність кожної системи може залежати від конкретних задач та вимог використання. Для того щоб вибрати найкращу систему для конкретного випадку, потрібно ретельно ознайомитись з можливостями та різними рішеннями.

Реалізація рішення

Після детального розгляду методів систем виявлення вторгнень було вирішено обрати *метод виявлення аномалій з* використанням технології машинного навчання. В комплексі ця система буде більш адаптивною та здатною реагувати як на старі, так і на нові, раніше невідомі, загрози.

Для навчання моделі було обрано набір даних від Канадського інституту кібербезпеки – CIC(*Canadian Institute of Cybersecurity*) IoT 2023, який був зібраний у реальному часі для масштабних атак у середовищі IoT. Це достатньо новий і розширений набір даних про атаки в IoT для сприяння розробці додатків, аналітик і безпеки.

В наборі відзначені дані з 33 атак, розділених на 7 класів (рис.1).

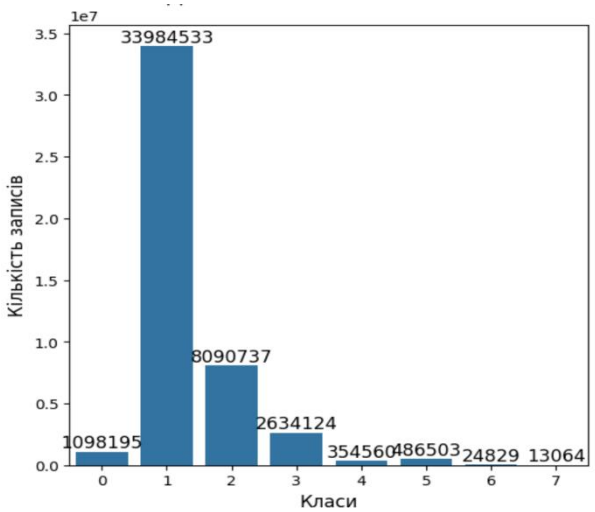


Рис. 1 Класи атак

За допомогою нього буде навчена нейронна мережа, яка буде виконуватися в системі виявлення вторгнень, яка є метою цього документу, щоб класифікувати та виявляти мережевий трафік IoT, як зловмисний або безпечний.

Робота з підготовки даних, навчання та тестування буде проводитись у середовищі *Jupyter Notebook* на мові *Python,* версії 3.11. Набір даних розділено на піднабори, тож напочатку роботи об'єднаємо їх в один великий, це зменшить продуктивність системи з точки зору пам'яті, але дасть нам мобільності при виконанні тих чи інших операцій в процесі навчання чи підготовки до навчання. Набір даних містить 46686579 записів, 46 ознак для навчання і клас-ознаку для класифікації (рис. 2).

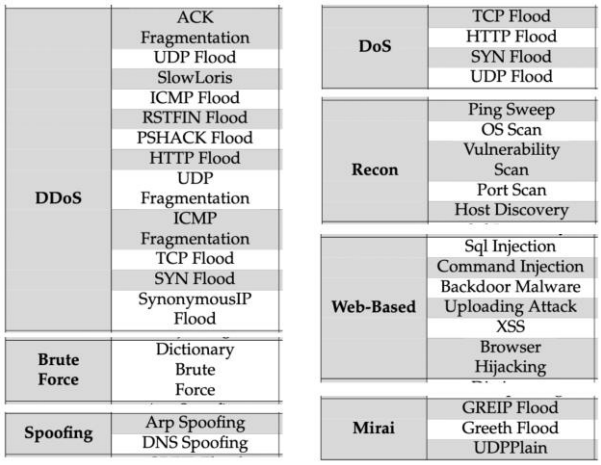


Рис. 2. Розподіл класів за кількістю записів:
0 – звичайний трафік, 1 – DDoS; 2 – DoS; 3 – Mirai;
4 – Recon; 5 – Spoofing; 6 – Web-Based, 7 – BruteForce

Переходимо до фази *підготовки даних,* яка є найважливішою у процесі машинного навчання, бо якість та обсяг даних безпосередньо впливають на результати роботи моделі. Очищаємо дані, видаляємо відсутні значення, скидаємо індекс нашого фрейму даних і використовуємо замість нього стандартний, видаляємо рядки, які повторюються.

Визначаються важливі ознаки(певні характеристики з набору даних), які далі будуть використовуватися для навчання моделі, нормалізуються дані, для забезпечення стабільності та швидкості навчання, розділяються на тренувальні та тестові набори. Правильна підготовка даних є критичним етапом, який може визначити невдачу чи успіх моделі в подальшому навчанні та роботою з реальними даними.

Вибір моделі ансамблевого методу і оптимізація параметрів

В ході досліджень було вирішено використовувати модель ансамблевого методу *RandomForest.* Це метод машинного навчання, який використовується для класифікації та регресії. Він є типом і відповідає ряду класифікаторів дерева рішень на різних підвибірках набору даних. Використовує техніку випадковості і усереднення для підвищення точності прогнозування, покращення продуктивності та стабільності моделі. *RandomForest* включає в себе кілька дерев рішень, кожне з яких навчається на випадковій підмножині даних та ознак. Коли треба прийняти рішення, модель об'єднує прогнози всіх дерев, зазвичай за допомогою класифікації або середнього значення для регресії. Модель має властивість стійкості до пере-

навчання, оскільки кожне дерево навчається на випадковій підмножині даних та ознак. Це дозволяє ансамблю підтримувати генералізацію на нових, раніше не бачених ознак [10].

Перед навчанням подбаємо про гіперпараметри та оптимізуємо їх за допомогою бібліотеки *optuna.* Для моделі *RandomForest* нас цікавлять:

max_depth – максимальна глибина кожного дерева в ансамблі, яка визначає кількість рівнів у дереві рішень.

max_features – визначає максимальну кількість ознак, які випадково обираються для розгляду при побудові кожного дерева в "лісі".

n_estimators – гіперпараметр який вказує кількість дерев, які мають бути побудовані в ансамблі.

Коли гіперпараметри визначені і оптимізацію завершено, починається навчання, яке буде займати деякий час. Для оцінки ефективності моделі використовуємо метрики: *Accuracy*, *Precision, Recall, F score,* які враховують різні показники результатів класифікацій та дозволяють отримати більш повну картинку продуктивності моделі. Розберемо більш детально кожний з них:

Accuracy (правильність) – частка прогнозів, яку наша модель отримала правильно. Математично це співвідношення між кількістю правильних пронозів до загальної кількості прогнозів. Це корисно, коли всі класи мають однакову важливість, як у нашому випадку але є недолік зі сторони незбалансованого набіру даних.

Precision (точність) - це співвідношення $\frac{TP}{TP+FP}$, де *TP* – кількість справжніх спрацьовувань, а *FP* – кількість хибних спрацьовувань [11]. Точність – це інтуїтивно зрозуміла здатність класифікатора не позначати негативний зразок як позитивний.

Recall (запам'ятовування) – це відношення $\frac{TP}{TP+FN}$, де *TP* – кількість справжніх позитивних результатів, а *FN* –помилкових негативних результатів.

З*апам'ятовування* – це інтуїтивно зрозуміла здатність класифікатора знаходити всі позитивні зразки [11].

F-оцінка -може бути інтерпретована як зважене гармонічне середнє значення точності та запам'ятовування, досягає найкращого значення при 1, а найгіршого при 0. Визначається як:

$$F_1 = \frac{2}{\frac{1}{recall} \times \frac{1}{precision}} = 2 \times \frac{precision \times recall}{precision + recall}.$$

За підсумком навчання маємо результати, наведені в табл. 1, 2 та рис. 3.

Таблиця 1 – **Кількісна оцінка якості**

	Precision	**Recall**	**F1-Score**	**Support**
0	0.91	0.98	0.94	878773
1	1.00	1.00	1.00	27188627
2	1.00	1.00	1.00	6470759
3	1.00	1.00	1.00	2107532
4	0.91	0.84	0.87	283896
5	0.92	0.86	0.89	389289
6	0.98	0.54	0.70	19893
7	0.99	0.57	0.72	10467
accuracy			1.00	37349236
macro avg	0.96	0.85	0.89	37349236
weighted avg	1.00	1.00	1.00	37349236

Таблиця 2 **– Метрики ефективності моделі**

Train Score	**Test Score**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
0.9974162791442	0.9963210760187	0.9963210760187	0.9646470264108	0.8473565625810	0.8903067510023

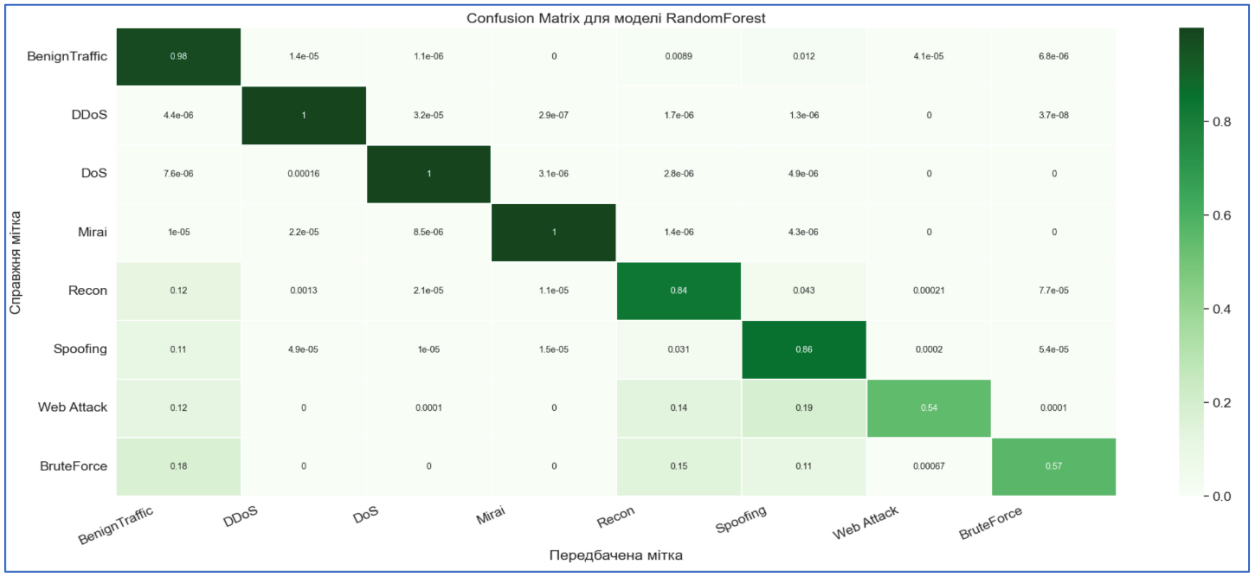


Рис. 3. Матриця помилок для моделі *Random Forest*

Після тестування модель зберігається локально за допомогою бібліотеки *pickle.* На виході модель *RandomForest,* це набір дерев-предикторів $\{t(x_{in}, \theta_n), n = 1, \dots\}$, які індивідуально роблять передбачення на заданому параметрі x_{in}. Кожен предиктор залежить від випадкового набору змінних $\{\theta_n\}$, які незалежно відбираються з однаковим розподілом (рис. 4).

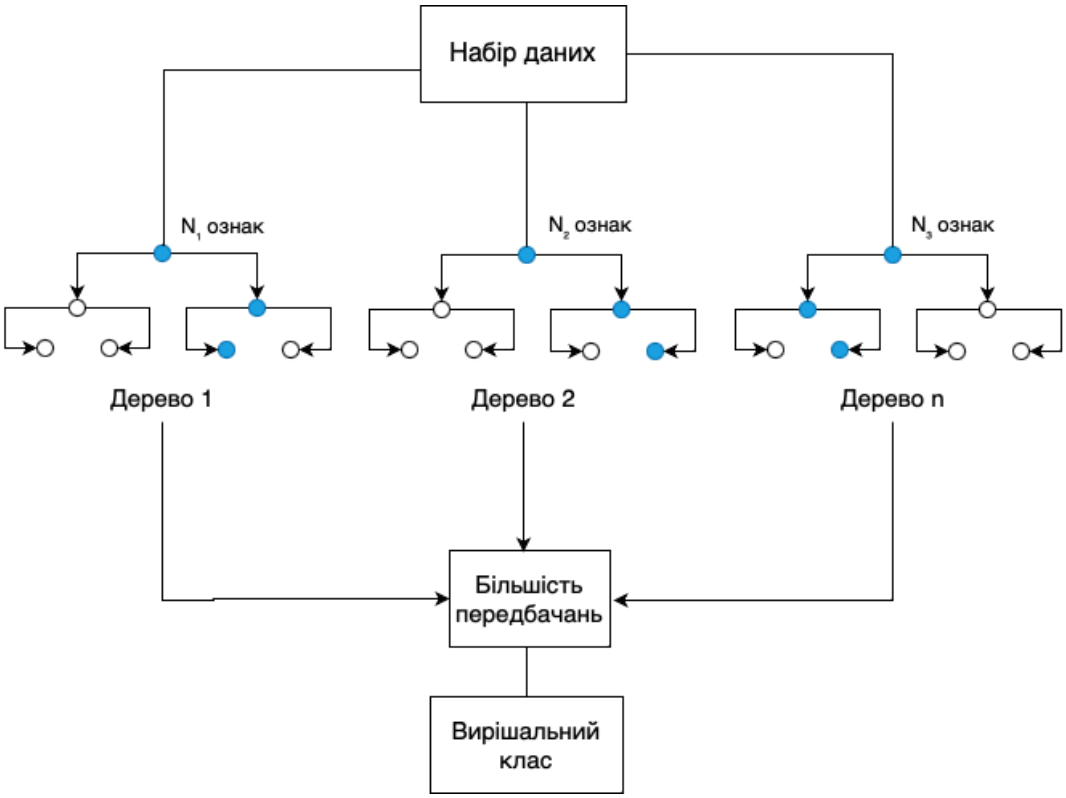


Рис. 4 Архітектура моделі *Random Forest*

Реалізація системи виявлення вторгнень з нейронною мережею (рис. 5)

Додаток буде в реальному часі зчитувати інтернет пакети або *.pcap* файли, діставати з них усі необхідні ознаки та відправляти моделі для отримання передбачення. За допомогою бібліотеки *pyshark,* для захоплення та аналізу інтернет-пакетів, витягуємо необхідні ознаки для подальшого використання. Після того як всі ознаки витягнуті відправляємо їх до попередньо навченої моделі для отримання передбачень. Обробляємо результати передбачень та приймаємо рішення щодо подальших дій, сповіщення або вжиття інших заходів безпеки. Було вирішено розробити інтерфейс командного рядка (*CLI – Command-line interface,* рис. 6). Користувач зможе встановити цей додаток за допомогою системи керування пакунками (*pip*) на операційну систему Windows або класу Linux. Він матиме змогу запускати його виконання з командного рядка або додати в автоматичний запуск за допомогою скриптів.

Рис. 5 Архітектура IDS

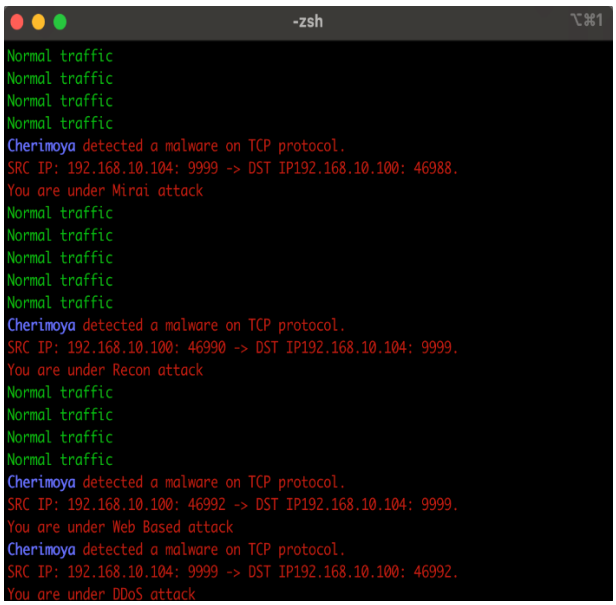


Рис. 6 Інтерфейс командного рядка (*CLI)* застосунку Cherimoya (*назва системи виявлення вторгнень)*

Висновки

У наслідок проведених досліджень і розглянутих концепцій передових систем та методів виявлення атак в інфраструктурі IoT було запропоновано власну систему виявлення вторгнень в реальному часі, яка базується на методі виявленні аномалій, та працює в комплексі з нейронної мережею ансамблевого методу *RandomForest,* яка була навчена на наборі з великим обсягом даних та має гарні показники:

правильність – 0.996;

точність – 0.964;

запам'ятовування – 0.847;

г*армонічне середнє значення точності та запам'ятовування – 0.89.*

Для зручності використання застосунку системи був розроблений інтерфейс командного рядка, котрий сповіщає користувача або іншу систему про вторгнення чи атаку. В майбутньому запропонована модель може бути використана для систем побудованих в поєднанні з концепцією *туманного обчислення* та Інтернету речей, на принципах Fog-IoT архітектури.

Список літератури

1. T. Mazhar, D. B. Talpur, T. Al Shloul, Y. Y. Ghadi, I. Haq, I. Ullah, K. Ouahada, H. Hamam Analysis of IoT Security Challenges and Its Solutions Using Artificial Intelligence. 2023. DOI: https://doi.org/10.3390%2Fbrainsci13040683
2. A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman Survey of intrusion detection systems: techniques, datasets and challenges. 2019. URL: https://cybersecurity.springeropen.com/articles/10.1186/s42400-019-0038-7
3. Рубан І. В. Класифікація методів виявлення аномалій в інформаційних системах / І. В. Рубан, В. О. Мартовицький, С. О. Партика // Системи озброєння і військова техніка. — 2016. — № 3. — С. 100-105
4. Verma Abhisher, Virender Ranga Machine learning based intrusion detection systems for IoT applications. 2020. URL: https://link.springer.com/article/10.1023/A:1010933404324
5. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об'єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: https://doi.org/10.20998/2522-9052.2018.1.04
6. Ruban, I.V., Martovytskyi, V.O., Kovalenko, A.A. and Lukova-Chuiko, N.V. (2019), "Identification in Informative Systems on the Basis of Users' Behaviour", Proceedings of the International Conference on Advanced Optoelectronics and Lasers, CAOL 2019-September,9019446, pp. 574-577, DOI: https://doi.org/10.1109/CAOL46282.2019.9019446
7. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), "Horizontal scaling method for a hyperconverged network", *2021 International Conference on Information and Digital Technologies* (IDT), Zilina, Slovakia, doi: https://doi.org/10.1109/IDT52577.2021.9497534
8. J. Delua Supervised vs. Unsupervised learning. 2021. URL: https://www.ibm.com/blog/supervised-vs-unsupervised-learning/
9. 1. I. U. Khan, M. Ouaissa, M. Ouaissa, Z. A. El Houda, M. Fazal Cyber Security for Next-Generation Computing. 2024. DOI: https://doi.org/10.1201/9781003404361
10. . Журило, О., Ляшенко, О. і Аветісова, К. 2023. ОГЛЯД РІШЕНЬ З АПАРАТНОЇ БЕЗПЕКИ КІНЦЕВИХ ПРИСТРОЇВ ТУМАННИХ ОБЧИСЛЕНЬ У ІНТЕРНЕТІ РЕЧЕЙ. СУЧАСНИЙ СТАН НАУКОВИХ ДОСЛІДЖЕНЬ ТА ТЕХНОЛОГІЙ В ПРОМИСЛОВОСТІ. 1 (23) (Квіт 2023), 57–71. DOI: https://doi.org/10.30837/ITSSI.2023.23.057.
11. V. Martovytskyi, I. Ruban, H. Lahutin, I. Ilina, V. Rykun and V. Diachenko, "Method of Detecting FDI Attacks on Smart Grid," 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 2020, pp. 132-136, doi: 10.1109/PICST51311.2020.9468005

Received (Надійшла) 05.11.2023
Accepted for publication (Прийнята до друку) 24.01.2024

Model and methods of detection of a large-scale attack in the IoT environment

Oleksii Liashenko, Ihor Velykodnyi, Vasyl Znaidiuk, Oleh Zhurylo

Abstract. The main concept and subject of the study is the detection of various types of extensive attacks in the IoT infrastructure, an overview of the presented model, methods and existing advanced intrusion detection systems. **The purpose of this work** is to propose a real-time intrusion detection system that will be trained on a large data set using a neural network using an ensemble machine learning **method**. **The subject of the research** is an overview of existing methods and models for detecting a large-scale attack and proposing an intrusion detection system solution, which will be based on the method of detecting anomalies and a neural network. **Conclusion**. An intrusion detection system was built, which analyzes Internet traffic, extracts signs from the packet, processes them and predicts various types of attacks, as well as characterizes them by type. Security threat can be considered as the main critical issue for IoT devices, so the use of such systems reduces the risks of data loss.

Keywords: dataset, neural network, machine learning, network traffic, IDS, training, prediction, anomaly detection, attack, Internet of Things, IoT, intrusion detection system.

УДК 004.7 **doi: 10.26906/SUNZ.2024.1.133**

Р. М. Марченко, А. А. Коваленко, В. Г. Знайдюк

Харківський національний університет радіоелектроніки, Харків, Україна

АНАЛІЗ МЕТОДІВ ВИЯВЛЕННЯ АНОМАЛЬНОГО ТРАФІКУ В МЕРЕЖАХ IoT

Анотація: Метою даної роботи є проведення комплексного аналізу методів та підходів до виявлення аномалій в мережах Інтернету речей (IoT). З урахуванням стрімкого розвитку IoT і збільшення кількості підключених пристроїв, проблема виявлення аномального трафіку стає актуальною для забезпечення безпеки та ефективності цих мереж. У роботі розглядаються різні методи та підходи до виявлення аномалій, включаючи статистичний аналіз, мережевий моніторинг, поведінковий аналіз, а також застосування сучасних технологій машинного та глибокого навчання. Кожен із цих методів розглядається з точки зору його застосовності в контексті IoT та оцінюються його переваги та обмеження. Робота також розглядає сучасні виклики і перспективи розвитку у галузі безпеки IoT, з фокусом на захисті від кіберзагроз та посиленні систем виявлення аномалій.

Ключові слова: Інтернет речей, аномальний трафік, машинне навчання, глибинне навчання, статистичний аналіз.

Вступ

Розвиток Інтернету речей (IoT) та пов’язаних технологій супроводжується експоненціальним зростанням кількості пристроїв різних типів, що працюють на основі різноманітних технологій та підключені до мережі, яка об'єднує їх локально та через мережу Інтернет. Характерною рисою IoT є збільшення кількості сенсорів, які збирають дані з навколишнього середовища, а потім аналізують ці дані та впливають на фізичний світ через виконавчі механізми.

Пристрої IoT використовуються у багатьох сферах, включаючи побутову техніку, охоронні системи, медичне обладнання, системи управління та носимі пристрої. За сучасними оцінками, кількість підключених до IoT пристроїв щодня зростає, і до 2025 року їх може бути приблизно 85 мільярдів, що охоплюватиме галузі виробництва (40%), медицину (30%), роздрібну торгівлю та безпеку (20%) [1].

Цей суттєвий розвиток IoT відкриває нові можливості для майбутніх застосувань. У міру збільшення цінності даних, що зберігаються, обробляються та передаються, разом із масштабом зростають і атаки на них [2–5]. Ці прогнози показують, що кількість і рівень загроз і атак на пристрої IoT зростатиме, що потребуватиме більш надійних заходів безпеки. Використання IoT охоплює різноманітні сценарії, від окремих пристроїв до розгортання технологій на крос-платформеному рівні та використання систем реального часу у хмарних обчисленнях [6].

Функціональність в мережі IoT включають три основні завдання: передачу даних, отримання даних та їх обробку. На рівні застосунків, вбудовані інтерфейсні модулі дозволяють пристроям взаємодіяти з основною архітектурою. План управління пристроями визначає джерело та призначення даних для забезпечення операцій введення-виведення у пристроїв. Наприклад, агрегатор об'єднує дані, надані різними пристроями, в єдиний набір. Шар зв'язку виступає проміжним рівнем з мережними компонентами, які встановлюють різні протоколи та стандарти для керування трафіком у системі.

Використання стандартних протоколів дозволяє реалізувати належну комунікацію між пристроями IoT. Для таких систем важливий наявний набір простих правил для ініціалізації та обміну інформацією. Схематично багаторівневу архітектуру IoT можна представити у вигляді рис. 1, де наведено структуровану архітектуру типової системи IoT з відокремленими рівнями апаратного забезпечення, комунікацій та застосунків.

На найнижчому рівні знаходяться фізичні IoT-пристрої, які відповідають за збір даних та взаємодію з фізичним світом. Над ним розташований комунікаційний рівень, який містить протоколи зв'язку для транспортування даних в мережі. Цей рівень також включає агрегаційний шар, що узагальнює дані з різних джерел перед подачею на верхні рівні. Найвищий рівень включає веб-портали, управління API та хмарні/граничні сервіси, що надають різноманітні сервіси для обробки та аналітики подій.

Окремо від основних рівнів зображено план управління пристроями, який забезпечує інтеграцію та координацію всіх IoT-пристроїв у мережі.

Мета статті – провести аналіз методів виявлення аномального трафіку в мережах IoT, виявити основні переваги і проблеми цих методів для подальшого їх дослідження та впровадження, а також вплив факторів на їх ефективність.

Аналіз сучасних методів виявлення аномалій в IoT

Аномалія в контексті IoT – це дані або спостереження, які виходять за межі очікуваної поведінки в системі. Це може бути рідкісна подія або відхилення від типового шаблону в конкретний момент часу або для певного контексту. Аномалії можуть бути спричинені зовнішніми факторами, такими як помилки датчиків або кібератаки. Задача алгоритму виявлення аномалій – виявити ці викиди і, за можливості, визначити їхні причини [7].

Алгоритми виявлення аномалій можна розділити на чотири категорії в залежності від підходу до: вирішення задачі; способу застосування; типу методу; затримки алгоритму. Важливо мати різні підходи для різних застосувань IoT через їхню різноманітну природу та різновиди даних. Наприклад, один підхід може бути кращим для виявлення аномалій у вимірах датчиків, а інший – для виявлення відхилень у мережному трафіку.

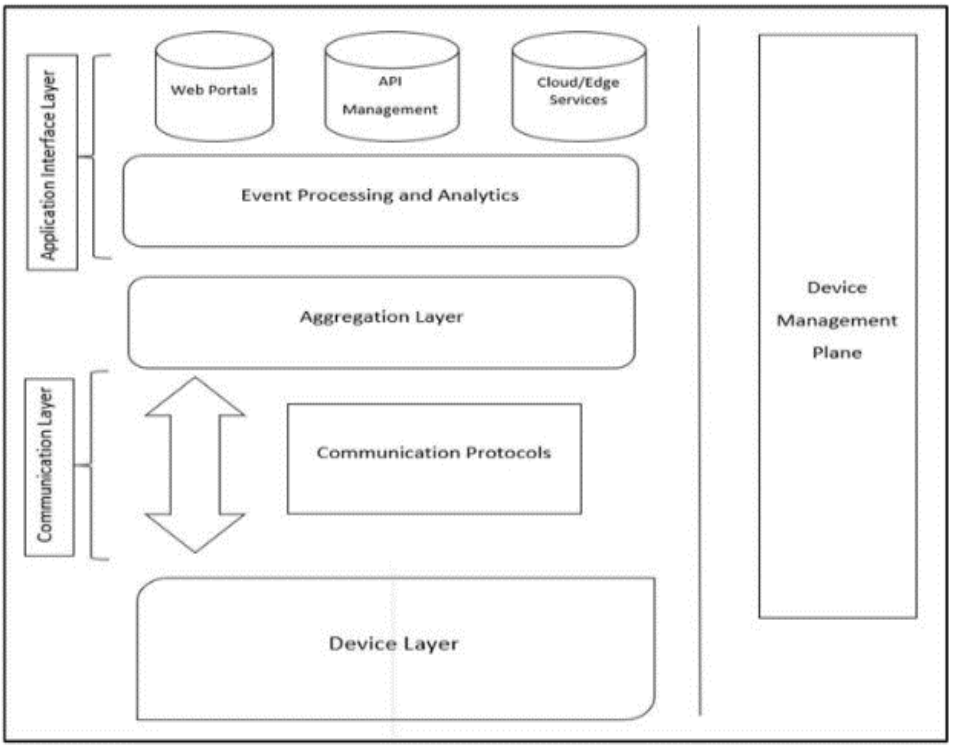


Рис. 1. Багатошарова архітектура IoT

У задачі бінарної класифікації аномалій велике значення має вибір моделі наближення, яка найкраще відображає очікувану поведінку даних. Точність цієї моделі визначає, наскільки ефективно будуть виявлені аномалії. Оскільки IoT включає в себе різноманітні застосунки та типи даних, часто потрібно використовувати різні стратегії для виявлення аномалій, які оптимізовані для конкретних сценаріїв. На рис. 2 показаний приклад однієї з аномалій [7].

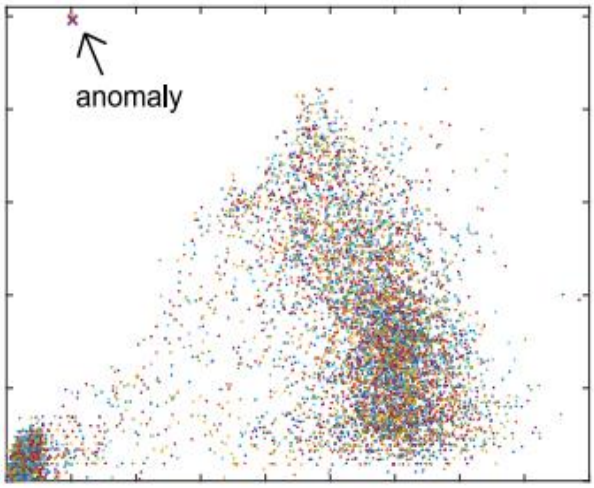


Рис. 2. Діаграма з прикладом аномалії

Методи виявлення аномалій в IoT поділяються на чотири категорії, комбінуючи класифікації з опублікованих результатів досліджень [8] та [9]. Їх класифікують за способом підходу до проблеми, застосуванням, типом методу та затримкою алгоритму. Нижче наведено короткий огляд цих методів та деяких традиційних підходів, що використовуються в IoT:

Класифікація за методом включає:

- геометричні методи: ґрунтуються на ідеї, що при стратегіях, основаних на відстані та щільності даних, очікувані та аномальні дані розділені; зазвичай вони використовують статичний або динамічний поріг для класифікації даних як нормальних або аномальних; декілька прикладів геометричних методів включають в себе методи на основі відстані та щільності;

- статистичні методи: намагаються моделювати нормальні дані за допомогою математичних моделей та розподілів; один із прикладів – метод мінімального об'єму, який намагається створити n-вимірний симплекс навколо заданої області даних;

- методи машинного навчання та глибокого навчання: вибір моделі залежить від характеру наданих даних; наприклад, моделі типу Long Short-Term Memory (LSTM) та трансформатори відповідають за послідовні дані, такі як аудіо, відео та часові ряди; з іншого боку, моделі типу Convolutional Neural Network (CNN) та Autoencoder (AE) підходять для не послідовних даних, таких як зображення.

Класифікація за застосуванням включає:

- конструктивні застосування: спрямовані на позитивну діяльність та надають користь, таку як моніторинг щоденної активності літніх людей для попередження падінь;

- деструктивні застосування: спрямовані на завдання шкоди, такі як атаки на мережу IoT або намагання завдати шкоди даним та застосункам;

- застосування для очищення даних: спрямовані на видалення непотрібних даних або шуму з вхідного сигналу.

Класифікація за типом аномалії включає:

- пунктові аномалії: виникають, коли одна точка даних відхиляється від очікуваної поведінки;

прикладом може бути виявлення шахрайства з банківськими картками;

- контекстуальні аномалії: аномалії, які можуть вважатися такими лише в певному контексті і виявляються, коли розглядаються як контекстуальні, так і поведінкові характеристики;

- колективні аномалії: визначаються на основі всього набору даних та не пов'язані з окремими точками даних.

Класифікація за затримкою включає:

- online алгоритми: обробляють дані під час їх збору і можуть аналізувати одну точку даних або вікно даних без повного доступу до всіх даних;

- offline алгоритми: мають доступ до всіх даних і використовують більш складні обчислювальні методи для розв'язання задачі.

Ця категоризація вказує на різноманітність методів виявлення аномалій в IoT та їх застосування в залежності від конкретного сценарію та потреби.

Основні переваги та проблеми методів виявлення аномалій в IoT, які вимагають подальшого дослідження наведені в табл. 1. Наведена таблиця результатів аналізу надає загальний огляд переваг та недоліків кожної категорії методів виявлення аномалій в Інтернеті речей залежно від різних аспектів їх використання та застосування.

Таблиця 1 – **Результати аналізу**

Категорія методу	Переваги	Недоліки
За методом		
Геометричні методи	Добре підходять для даних з чітко визначеними структурами.	Можуть бути неефективними для даних із складними структурами або часово залежними даними.
Статистичні методи	Можуть моделювати різноманітні розподіли даних.	Вимагають чіткого розуміння розподілу даних, що моделюються, і можуть бути неефективними для даних зі складними структурами або змінними з часом
Методи машинного навчання та глибокого навчання	Можуть виявляти складні аномалії та залежності між даними.	Вимагають великої кількості даних для тренування. Можуть бути складними для налаштування та оптимізації.
За застосуванням		
Конструктивні застосування	Надають користь та вирішують практичні завдання.	Вимагають розробки специфічних застосунків для кожного випадку.
Деструктивні застосування	Допомагають виявляти та запобігати шкідливим діям та атакам.	Зазвичай потребують додаткових заходів для захисту системи. Можуть призводити до фальсифікації або неправильного реагування.
Застосування для очищення даних	Допомагають видалити непотрібні дані та шум з даних.	Можуть втрачати корисну інформацію. Вимагають заздалегідь відомих шаблонів для очищення.
За типом аномалії		
Пунктові аномалії	Відокремлюють аномалії, які виникають в окремих точках даних.	Можуть пропустити аномалії, які виникають лише в контексті.
Контекстуальні аномалії	Враховують контекст та поведінкові характеристики для виявлення аномалій.	Вимагають складніших аналітичних методів та більше обчислювальних ресурсів.
Колективні аномалії	Визначають аномалії на основі всього набору даних та структури взаємозв'язків між даними.	Можуть бути обчислювально витратними та вимагати великої кількості даних для навчання.
За затримкою		
Online алгоритми	Здатні обробляти дані під час їх збору та аналізувати їх в реальному часі.	Можуть бути обмеженими за ресурсами та вимагати низької затримки.
Offline алгоритми	Мають доступ до всього набору даних і можуть використовувати більш складні обчислювальні методи.	Зазвичай вимагають більше обчислювальних ресурсів та можуть бути повільнішими в роботі.

Обмеження та вимоги до методів виявлення аномалій в IoT

Методи виявлення аномалій включають в себе етап попередньої обробки для визначення нормального діапазону значень, де будь-яке значення в межах визначеного діапазону вважається нормальним. Натомість будь-яке інше значення є винятком. Для потоку даних, залежного від часу, стандартний діапазон значень може змінюватися в залежності від повторюваного циклу, такого як сезон або різні повторювані часові інтервали. Тому правильне визначення повторюваного циклу має вирішальне значення для точності процесу виявлення аномалій.

Отже, слід зазначити наступні важливі обмеження та вразливості:

- визначення довжини повторюваного циклу є найважливішим кроком в аналізі даних IoT; неправильна довжина циклу призводить до невірного виявлення аномалій;

- виявлення аномалій на початку та в кінці кожного циклу є більш складним, оскільки різниця між нормальним станом та аномальним станом є незначною; отже, ймовірність помилки є значною;

- для підтримки точності в стандартних показниках, вимагається постійно перевіряти правильність значень оболонок та їх адаптацію до визначеного циклу і передбачати природні та обґрунтовані зміни в

циклі та відповідні значення, що використовуються для перевірки аномалій з плином часу.[1]

Окрім, того з огляду на результати аналізу, що наведено у попередніх підрозділах, можна сформулювати додаткові обмеження та вимоги до методів виявлення аномалій в IoT. Висока точність: методи виявлення аномалій повинні бути досить точними у виявленні незвичайних подій або аномалій. Особливо важливо виявляти аномалії в реальному часі для запобігання можливим проблемам. Адаптованість до змін: середовище IoT може змінюватися, і методи повинні бути адаптованими до нових умов та типів даних. Вони повинні бути здатними навчатися на нових даних та оновлювати моделі.

Низька обчислювальна складність: оскільки IoT може включати велику кількість пристроїв з обмеженими ресурсами, методи повинні бути ефективними з точки зору обчислень і споживання енергії.

Здатність до роботи в режимі реального часу: деякі випадки виявлення аномалій вимагають негайного реагування. Методи повинні бути здатними працювати в режимі реального часу та виявляти аномалії негайно.

Робота з різними типами даних: IoT може генерувати різноманітні типи даних, від сенсорних даних до великих обсягів текстової інформації. Методи повинні бути придатними для роботи з різними видами даних.

Захист від фальсифікації та атак: методи повинні бути відповідними до заходів з безпеки, оскільки IoT може бути піддається атакам та фальсифікації даних.

Масштабованість: методи повинні бути придатними для роботи в масштабах, що відповідають IoT, де кількість пристроїв і обсяги даних можуть бути дуже великими.

Висновки

У даній статті розглянуто основні методи виявлення аномалій в мережах IoT, проаналізувавши їх ключові переваги та недоліки. Розглянуто методи та підходи, включаючи геометричний, статистичний методи та методи машинного та глибокого навчання, зі спеціальним акцентом на їх застосовність в контексті IoT.

Напрямком подальших досліджень є пошук способу оптимізації існуючих методів виявлення аномалій в мережах IoT. Це включає пошук нових стратегій для підвищення точності та надійності методів виявлення, а також зниження впливу помилкових спрацьовувань.

СПИСОК ЛІТЕРАТУРИ

1. Parimala, V. K. (Ed.). (2024). Anomaly Detection - Recent Advances, AI and ML Perspectives and Applications.IntechOpen.DOI: 10.5772/intechopen.110988. ISBN: 978-1-83769-027-5.
2. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об’єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: https://doi.org/10.20998/2522-9052.2018.1.04
3. Ruban, I.V., Martovytskyi, V.O., Kovalenko, A.A. and Lukova-Chuiko, N.V. (2019), “Identification in Informative Systems on the Basis of Users' Behaviour”, Proceedings of the International Conference on Advanced Optoelectronics and Lasers, CAOL 2019-September,9019446, pp. 574-577, DOI: https://doi.org/10.1109/CAOL46282.2019.9019446
4. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), “Horizontal scaling method for a hyperconverged network”, *2021 International Conference on Information and Digital Technologies* (IDT), Zilina, Slovakia, doi: https://doi.org/10.1109/IDT52577.2021.9497534
5. Kuchuk, N., Kovalenko, A., Ruban, I., Shyshatskyi, A., Zakovorotnyi, O. And Sheviakov, I. (2023), “Traffic Modeling for the Industrial Internet of NanoThings”, *2023 IEEE 4th KhPI Week on Advanced Technology, KhPI Week 2023 - Conference Proceedings*, 194480, doi: http://dx.doi.org/10.1109/KhPIWeek61412.2023.10312856
6. Li, H., Boulanger, P. A Survey of Heart Anomaly Detection Using Ambulatory Electrocardiogram (ECG). Sensors. 2020; 20(5): 1461. DOI: 10.3390/s20051461.
7. Cook, A. A., Mısırlı, G., & Fan, Z. (2020). Anomaly Detection for IoT Time-Series Data: A Survey. IEEE Internet of Things Journal, 7(7), 6481–6494.
8. M. Fahim, A. Sillitti, Anomaly detection, analysis and prediction techniques in IoT environment: A systematic literature review, IEEE Access 7 (2019) 81664–81681
9. Chatterjee, A., & Ahmed, B. S. (2022). IoT Anomaly Detection Methods and Applications: A Survey. Internet of Things, 100568. Elsevier BV.

Received (Надійшла) 29.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Analysis of Methods for Detecting Anomalous Traffic in IoT Networks

Roman Marchenko, Andriy Kovalenko, Vasyl Znaidiuk

Abstract: The aim of this work is to conduct a comprehensive analysis of methods and approaches for anomaly detection in Internet of Things (IoT) networks. Considering the rapid development of IoT and the increasing number of connected devices, the problem of detecting anomalous traffic becomes crucial for ensuring the security and efficiency of these networks. This study examines various methods and approaches to anomaly detection, including statistical analysis, network monitoring, behavioral analysis, as well as the application of modern machine learning and deep learning technologies. Each of these methods is considered from the perspective of its applicability in the context of IoT and its advantages and limitations are evaluated. The work also explores current challenges and future prospects in the field of IoT security, with a focus on protection against cyber threats and the enhancement of anomaly detection systems.

Keywords: Internet of Things, anomalous traffic, machine learning, deep learning, statistical analysis.

UDC 004.7 **doi: 10.26906/SUNZ.2024.1.137**

Oleksandr Mozhaiev, Nina Kuchuk, Demian Shtepa, Bohdan Sorobei

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

STUDY OF THE INTERNET OF THINGS NETWORK CONSTRUCTION TASKS

Abstract. The emergence and development of the Internet of Things stimulates the development of telecommunications and computing technologies. The current state of this process and its prospects allow the inclusion of a large number of devices connected to communication networks. This leads to the need to select an adequate model and methodological apparatus that allows working with such a quantity. The article discusses the concept of the IoT, which determines that any devices or things can now interact with each other at any time at any point in space.

Keywords: computer system, distributed system, fog computing, cloud computing, Internet of Things.

Introduction

Increasing the manufacturability of devices for receiving and transmitting information, reducing their cost and bringing together the capabilities provided by modern technologies and needs, due to the current state of social relations and human activity, served as an impetus for the development of the Internet of Things (IoT, Internet of Things). According to authoritative analysts, the total number of Internet of Things devices connected to communication networks is growing steadily. It has already exceeded the number of inhabitants on Earth.

In Fig. 1 shows statistical data for 2023 and a forecast for 2030 according to Statista [1, 2]. By 2030, the number of IoT devices is expected to be around 30 billion, approximately four times the human population. There are other forecasts.

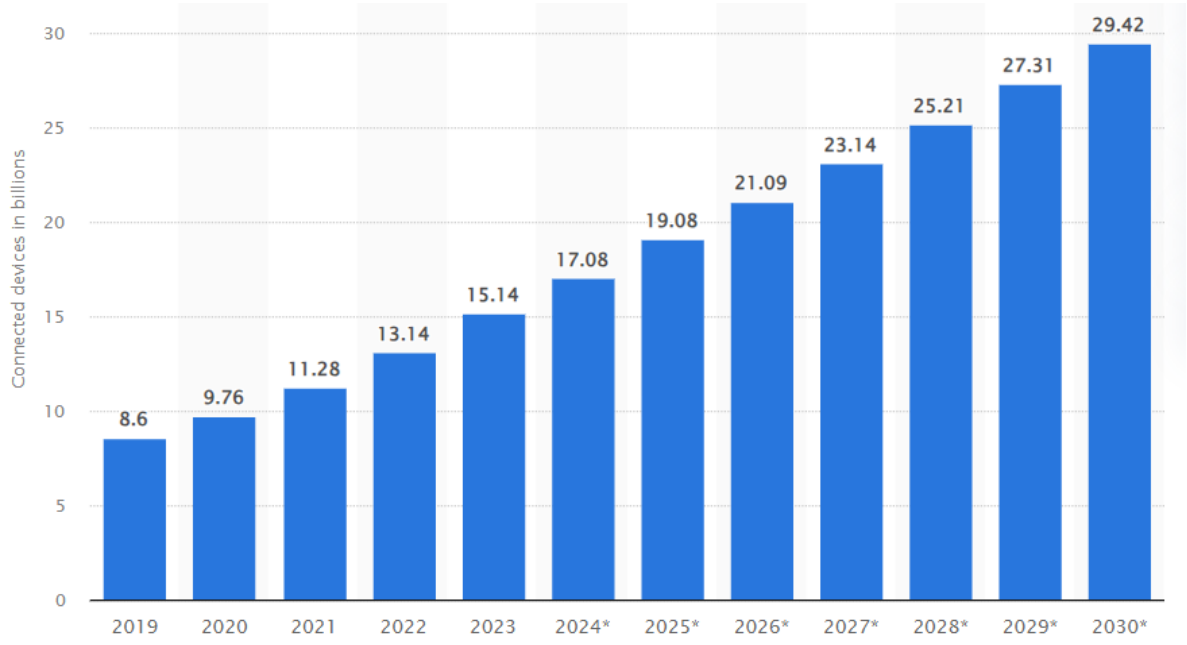


Fig. 1. Statistics and forecast of the growth in the number of connections to the number of IoT devices [1]

Main part

As defined by MCE-T in MCE-T-T Y.2060 [3, 4], the Internet of Things is “a global infrastructure for the information society that enables the provision of more complex services by combining (physical and virtual) things together based on existing and interoperable information and communication technologies (ICT).” The purpose of creating this infrastructure is to increase the availability of information in a global sense, which is illustrated in the Y.2060 recommendation [5, 6] as shown in Fig. 2.

The concept of this infrastructure assumes the presence of communication anywhere (on the street, at home, near a computer), at any time (day, night, moving) and any devices (between computers or devices between people) and devices directly between people without computers).

So, there is no doubt that IoT will develop in the long term. The main points of this process are [7, 8]:

• an increase in the number of devices connected to the interconnection in order to achieve a high-strength interconnection;

• the penetration of IoT technologies into various spheres of human activity, which entails the development of technical capabilities of these devices and advancements that are possible until the display of functional measures based on them;

• creation of both local specialized networks and penetration of this technology into global networks.

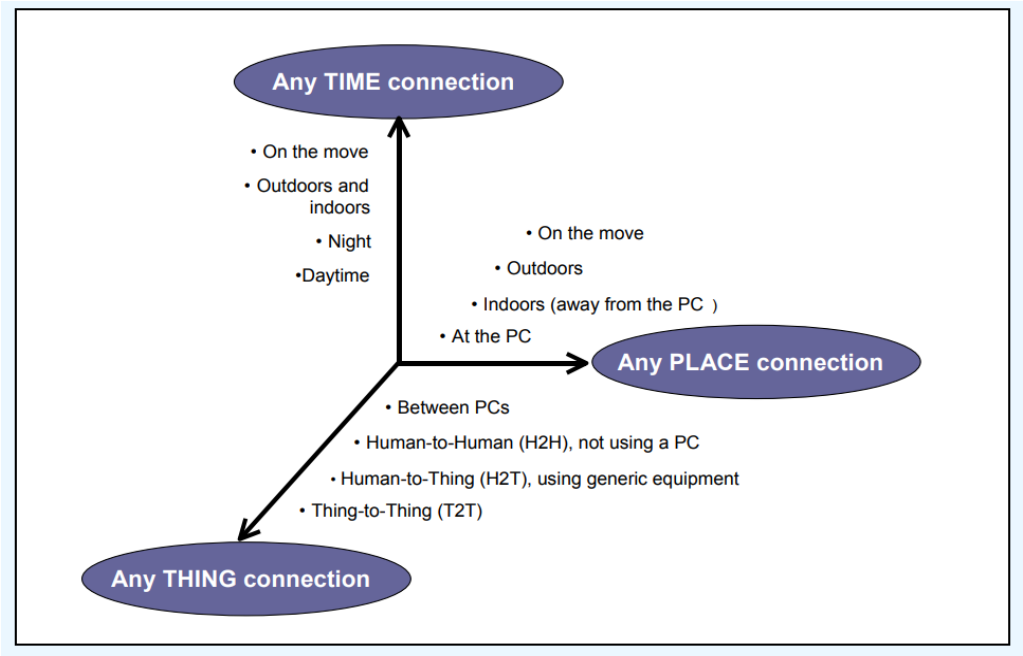


Fig. 2. Concept of the Internet of Things [2]

As a matter of greatest importance, it is important to consider the number of connections of devices, which outweighs the number of connections that were previously necessary in the boundaries, where the number of connections was assigned to Only a number of subscribers (customers).

The concept of the Internet of Speeches is in line with the recommendations of MCE-T and has low over-performing characteristics.

The growth of IoT leads to an increase in the number of devices connected to the limit. They are important for those that these devices are often localized in a whole area of space, so they should be brought to such a high concentration, so that the number of expansion per space increases. In this case, there may be a lack of flat (two-dimensional) model, the fragments of the device can be placed and interact in tridimensional space, and the third dimension (height) may be even more noticeable from the point of view organizing interaction between nodes. For example, a heterogeneous IoT network can include elements located on earthly platforms, including high-altitude and space platforms. Such gaps also occur in various areas: many surface water bodies and water bodies, at industrial sites and, possibly, in other situations. The main sign of this is that it is necessary to see how trivial the division of its nodes is in space. Since the service area is such that its “height” can be equalized with the other two dimensions, then such a measure can be considered as trivial, for example, in the area of altitude oblivion indicated in Fig. 3 [9, 10].

The most characteristic feature of a high-strength mesh is that in the area of connection of a sufficiently taken mesh node, it may appear that there are a lot of mesh nodes that actively flow into its work. This influx manifests itself in a reduced throughput capacity of the barrier. In fact, this imposes interconnection and can negatively affect the intensity of the IoT traffic generated by the nodes.

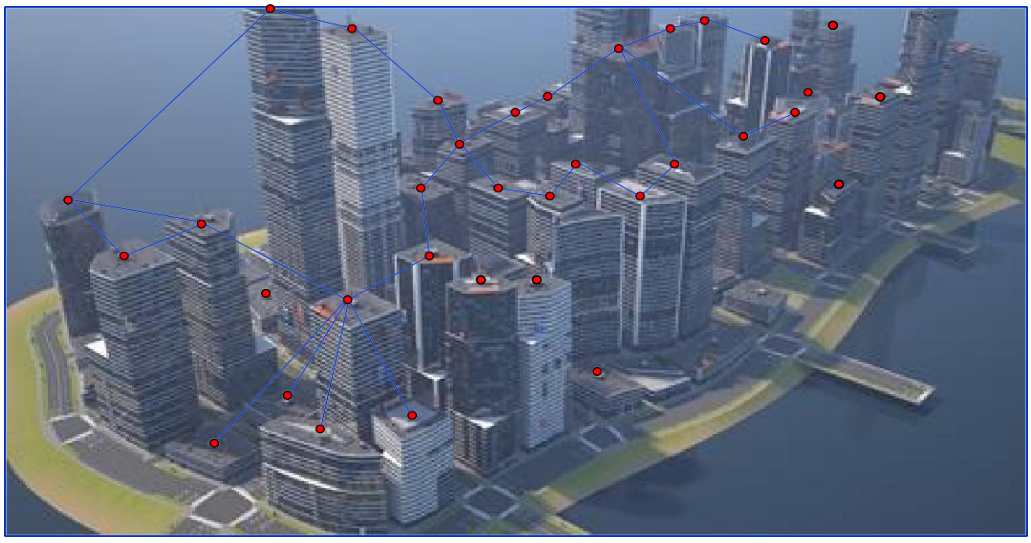


Fig. 3. Model of a potential site for hosting 3D networks of the Internet of Things [4]

Along with the negative properties, there are also positive qualities if the Internet of Things networks have a high density. Due to the high density of nodes, there is a high probability of finding a node near an arbitrarily chosen point in space [11]. This quality allows you to build a network of the structure that is necessary for some reasons, for example, it is one of the structures shown in Fig. 4.

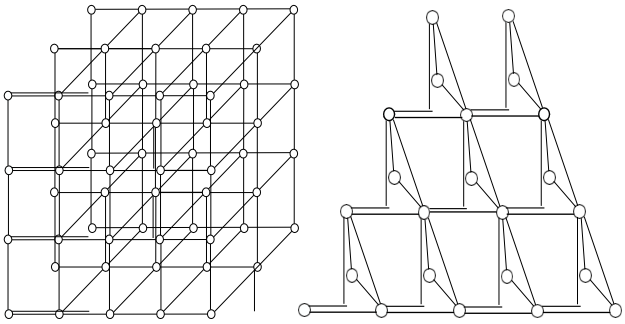


Fig. 4. Examples of high-density network structure modeling

This property also allows you to simplify some procedures. For example, in some cases it may be easier to find the shortest route in such a network, because the shortest distance is determined by a straight line, so the nodes included in this route are also most likely to be located near this straight line. Of course, such simplifications are not always possible. It is not always possible to obtain data about the coordinates of network nodes, so it is necessary to have methods that allow you to choose the network structure and manage it in conditions of a large number of nodes [12–14].

A three-dimensional network can have a different structure, in particular, it can be a regular structure in the form of a different kind of lattice or an arbitrary structure. However, in most cases, the structure of such a network is related to the structure of the environment in which it is created [15–17].

Selection of cluster master nodes in high-density Internet of Things networks. Let us assume that the Internet of Things network consists of *n* nodes:

$$V = \{v_1, \ldots, v_n\}, \quad (1)$$

distributed in three-dimensional space, an example is shown in Fig. 5.

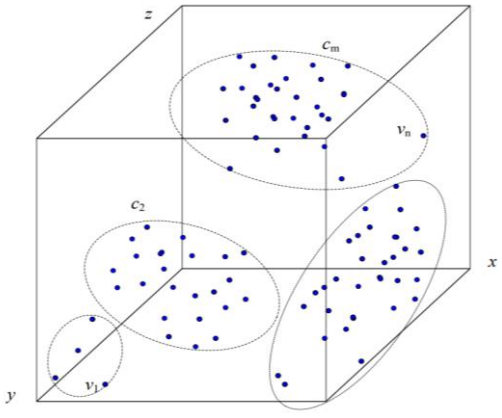


Fig. 5. Model of the Internet of Things network in three-dimensional space

We will also assume that clusters of nodes can be allocated in the network using the clustering method.

The task of building a network is to choose the positions of the main nodes in the Internet of Things network. In the general case, there may be several clusters in the network, making up a set of clusters

$$C = \{c_1, \ldots, c_K\} \quad (2)$$

from K extraordinary clusters, which can have an arbitrary shape and number of elements.

If nodes within a cluster form a connected ad-hoc network, then one or more master nodes can be used to service them, Fig. 6.

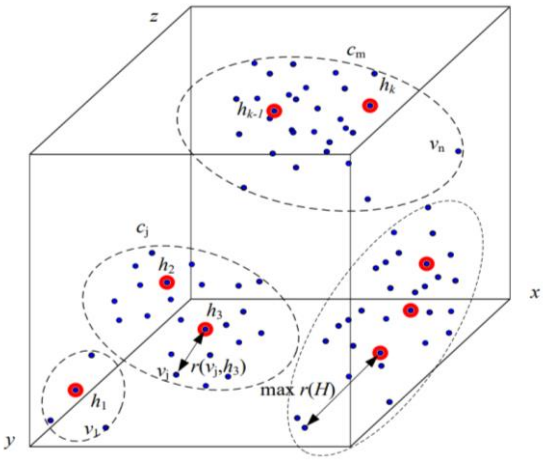


Fig. 6. Examples of a three-dimensional cluster

Let $r\,(v_j,h)$ be the distance between the *j-th* cluster node and the nearest master node. The position of the head node h should be chosen so as to minimize the distance to the most distant network node, i.e. ensure acceptable communication quality in the worst case.

This problem can be considered as an optimization problem in which the objective function

$$r(h) = \min\left(\max_{v \in V}\left(d(v,h)\right)\right), \quad h \in V. \quad (3)$$

It should be noted that selecting one node is often not enough; in this case, we are talking about selecting many main nodes.

$$H = \{h_1, \ldots h_k\}, \quad H \subset V. \quad (4)$$

Problem (4) is the task of finding the center of a graph formed by cluster nodes.

The center of the graph is the vertex from which the maximum distance (length of the shortest path) to the other vertices is minimal.

The problem of finding the center of a graph is solved by searching for all shortest paths in the graph, for example, using the Floyd-Warshall algorithm and finding the center of the graph from this data. Using this approach is quite acceptable if the number of cluster nodes (vertices in the graph) is not too large, both in terms of computational complexity and in terms of the ability to service the traffic they produce.

The computational complexity is that searching for all shortest paths requires time determined by the cube of the number of nodes (vertices) $O(n^3)$. This is a solvable

problem, but with a large number of nodes, the time it takes to solve it can be unacceptably long.

In the case of a high-density Internet of Things network, one can resort to simplifying the problem due to the fact that the length of the shortest path in such a network is close to the distance between vertices. This logically follows from the fact that the shortest distance between two points is determined by a straight line passing through them, and in the case of a high-density Internet of Things network, there is a high probability that the transit nodes of the path will be very close to this straight line. In this case, it is enough to know the distance between the nodes or their coordinates.

Serviceability is determined by the performance of the head node and the capacity of the routes. More often than not, one head node is not enough to serve all network nodes. Then the problem can be considered as a search for several nodes. This problem is known as the k-fold graph center problem.

Conclusions

So, to build high-density Internet of Things networks, it is necessary to develop a method that allows solving problems of modeling and managing a network with a large number of devices, that is, it takes into account the specific features of a given network.

REFERENCES

1. She R., Sun M. Security Energy Efficiency Analysis of CR-NOMA Enabled IoT Systems for Edge-cloud Environment. Int. Journal of Computational Intelligence Systems. 2023. Vol. 16(1), 118. DOI: http://dx.doi.org/10.1007/s44196-023-00273-y.
2. Kuchuk G., Nechausov S., Kharchenko, V. Two-stage optimization of resource allocation for hybrid cloud data store. *Int. Conf. on Information and Digital Technologies*. Zilina, 2015. P. 266-271. DOI: http://dx.doi.org/10.1109/DT.2015.7222982.
3. Кучук Г.А., Коваленко А. А., Лукова-Чуйко Н. В. Метод мінімізації середньої затримки пакетів у віртуальних з'єднаннях мережі підтримки хмарного сервісу. *Системи управління, навігації та зв'язку*. Полтава . ПНТУ, 2017. Вип. 2(42). С. 117-120.
4. Sharma, M., Kaur, P. Reliable federated learning in a cloud-fog-IoT environment. *Journal of Supercomputing*. 2023. Vol. 79(14). P. 15435–15458. DOI: http://dx.doi.org/10.1007/s11227-023-05252-w.
5. Baucas, M.J., Spachos, P. Improving Remote Patient Monitoring Systems Using a Fog-Based IoT Platform with Speech Recognition. 2023. *IEEE Sensors Journal*. Vol. 23(15). P. 17611–17618. DOI: http://dx.doi.org/10.1109/JSEN.2023.3287916.
6. Number of Internet of Things (IoT) connected devices worldwide from 2019 to 2023, with forecasts from 2022 to 2030. Statista. Telecommunications. URL: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide.
7. Kowalczyk A. European IoT Spending to Reach Nearly $227 Billion in 2023, Despite Ongoing Market Uncertainty, 2023. URL: https://www.idc.com/getdoc.jsp?containerId=prEUR250941023&utm_medium=embedd&utm_campaign=idc_embedd&utm_source=referral
8. Rehan M.M., Rehmani M.H. Blockchain-enabled Fog and Edge Computing: Concepts, Architectures and Applications: Concepts, Architectures and Applications. Taylor and Fransis, 2020. 302 p.
9. Jonathan Bar-Magen Numhauser. Fog Computing- Introduction to a new Cloud evolution. Proceedings from the CIES III Congress, January 2012 (англ.) // Escrituras silenciadas: paisaje como historiografía / José Francisco Forniés Casals (ed. lit.), Paulina Numhauser (ed. lit.), Proceedings from the CIES III Congress, January 2012.
10. Hamid Reza Arkian, Abolfazl Diyanat, Atefe Pourkhalili. MIST: Fog-based data analytics scheme with cost-efficient resource provisioning for IoT crowdsensing applications // Journal of Network and Computer Applications. – 2017-03-15. – Vol. 82. – P. 152–165. – ISSN 1084-8045. doi: http://doi.org/10.1016/j.jnca.2017.01.012
11. Kuchuk G., Kovalenko A., Komari I.E., Svyrydov A., Kharchenko V. Improving big data centers energy efficiency: Traffic based model and method. Studies in Systems, Decision and Control, vol 171. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (Eds.). Springer Nature Switzerland AG, 2019. Pp. 161-183. DOI: http://doi.org/10.1007/978-3-030-00253-4_8
12. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об'єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: https://doi.org/10.20998/2522-9052.2018.1.04
13. Кучук Н. Г., Мерлак В. Ю., Скородєлов В. В. Метод зменшення часу доступу до слабкоструктурованих даних. *Сучасні інформаційні системи*. 2020. Т. 4, № 1. С. 97-102. doi: https://doi.org/10.20998/2522-9052.2020.1.14
14. She R., Sun M. Security Energy Efficiency Analysis of CR-NOMA Enabled IoT Systems for Edge-cloud Environment. *Int. Journal of Computational Intelligence Systems*. 2023. Vol. 16(1), 118. DOI: http://dx.doi.org/10.1007/s44196-023-00273-y.
15. Петровська І. Ю., Кучук Г. А. Розподіл обчислювальних ресурсів у хмарних системах. *Системи управління, навігації та зв'язку*. 2022. Вип. 2 (68). С. 75–78. DOI: http://dx.doi.org/10.26906/SUNZ.2022.2.075.
16. Kuchuk G., Nechausov S., Kharchenko, V. Two-stage optimization of resource allocation for hybrid cloud data store. *Int. Conf. on Information and Digital Technologies*. Zilina, 2015. P. 266-271. DOI: http://dx.doi.org/10.1109/DT.2015.7222982.
17. Essalhi, S.E., Raiss El Fenni, M., Chafnaji, H. A new clustering-based optimised energy approach for fog-enabled IoT networks. *IET Networks*. Vol. 12(4). P.155–166. DOI: http://dx.doi.org/10.1049/ntw2.12082.

Received (Надійшла) 29.11.2023
Accepted for publication (Прийнята до друку) 24.01.2024

Вивчення задач побудови мережі Інтернету речей

О. О. Можаєв, Н. Г. Кучук, Д. С. Штепа, Б. В. Соробей

Анотація. Поява та розвиток Інтернету речей стимулює розвиток телекомунікаційних та обчислювальних технологій. Сучасний стан цього процесу та його перспективи дозволяють охопити велику кількість пристроїв, підключених до мереж зв'язку. Це призводить до необхідності вибору адекватної моделі та методичного апарату, що дозволяє працювати з такою величиною. У статті розглядається концепція IoT, яка визначає, що будь-які пристрої або речі тепер можуть взаємодіяти один з одним у будь-який час у будь-якій точці простору. для побудови мереж Інтернету речей високої щільності необхідно розробити метод, який дозволяє вирішувати задачі моделювання та управління мережею з великою кількістю пристроїв, тобто враховує особливості даної мережі.

Ключові слова: комп'ютерна система, розподілена система, туманні обчислення, хмарні обчислення, Інтернет речей.

УДК 004.8:007.5:159.9:303.77 **doi: 10.26906/SUNZ.2024.1.141**

В. В. Нарожний, В. С. Харченко

Національний аерокосмічний університет “Харківський авіаційний інститут”, Харків, Україна

МЕТОД СЕМАНТИЧНОГО АНАЛІЗУ ДАНИХ ДЛЯ ВИЗНАЧЕННЯ МАРКЕРНИХ СЛІВ ПРИ ОБРОБЛЕННІ РЕЗУЛЬТАТІВ ОЦІНКИ ВІЗИТОРІВ В ІНТЕРАКТИВНОМУ МИСТЕЦТВІ

Анотація. Предметом дослідження є поглиблений семантичний аналіз даних, що базується на інтеграції методологій латентного розподілу Діріхле (LDA) та двонаправленого кодувального представлення з трансформаторів (BERT). Це дослідження зосереджується на обробленні текстових даних, зокрема, оцінок відвідувачів інтерактивного мистецтва, для визначення слів-маркерів, які виділяють ключові емоційні та тематичні елементи. Мета: поглибити розуміння досвіду та сприйняття відвідувачами інтерактивних мистецьких інсталяцій шляхом визначення значущих слів-маркерів за допомогою комбінованого підходу LDA та BERT. Це комплексування має на меті охопити як загальний тематичний зміст, так і нюансований контекст зворотного зв'язку. Завдання: збір та попередня обробка текстових даних – оцінок відвідувачів, що складається з етапів токенізації, нормалізації та лематизації з впровадження LDA для виокремлення поширених тем із зібраних даних, що надає уявлення про основні теми, присутні у відгуках відвідувачів; інтеграція BERT для аналізу контекстуальних нюансів і виведення глибших значень з окремих слів у відгуках; поєднання результатів LDA та BERT для створення комплексного розуміння текстових даних, фокусуючись на виявленні найбільш значущих слів-маркерів. Досягнуто такі результати: виконано успішне виокремлення ключових тем з оцінок відвідувачів за допомогою LDA, що дозволило виявити широкі тематичні категорії, присутні у відгуках; запропоновано підхід глибокого навчання BERT, що забезпечив нюансовані контекстні вбудовування, підкреслюючи конкретні емоції та настрої, висловлені відвідувачами; здійснено інтеграцію результатів LDA та BERT, що надало багатий набір слів-маркерів, які ефективно відображають суть досвіду та сприйняття відвідувачами інтерактивного мистецтва; покращено точність і глибина аналізу у визначенні ключових емоційних і тематичних елементів, про що свідчить узгодженість і релевантність слів-маркерів відносно оцінок відвідувачів. Висновки: інтеграція LDA та BERT для семантичного аналізу даних в інтерактивних мистецьких контекстах демонструє потужний підхід для розуміння складних відгуків відвідувачів. Цей метод забезпечує дворівневий аналіз, де LDA пропонує розуміння загальних тем, а BERT сприяє детальному контекстуальному розумінню. Дослідження успішно визначає конкретні слова-маркери, які ефективно передають суть вражень та оцінок відвідувачів. Ця методологія може бути корисною для художників, кураторів та дослідників у вимірюванні публічної рецепції та покращенні інтерактивного мистецького досвіду. Адаптивність методології створює реальні перспективи її застосування в інших сферах, де потрібен детальний семантичний аналіз текстових відгуків.

Ключові слова: семантичний аналіз даних, обробка природної мови, прихований розподіл Діріхле, двонаправлені кодерні представлення з трансформаторів, інтерактивне мистецтво, аналіз емоційної реакції.

Вступ. Підхід до аналізу вербальних відгуків

Останніми роками розвиток цифрових платформ і соціальних мереж полегшив збір величезної кількості текстових відгуків від відвідувачів. Однак суб'єктивний і часто складний характер цих відгуків створює виклик: як ми можемо отримати значущі висновки з таких різноманітних і нюансованих даних? Саме тут відбувається перетин семантичного аналізу даних та обробки природної мови (NLP).

Семантичний аналіз даних передбачає вивчення тексту для виявлення закономірностей, тем і настроїв, які можуть бути неочевидними з першого погляду. Він дозволяє витягти глибші значення і зв'язки в тексті, забезпечуючи більш тонке розуміння змісту. У цьому дослідженні для вирішення цього завдання використано два інструменти NLP - латентний розподіл Діріхле (LDA) та двонаправлене кодування за допомогою трансформаторів (BERT).

LDA, популярна техніка тематичного моделювання, використовується для виявлення спільних тем і напрямків у зібраних відгуках. Він допомагає класифікувати текстові дані за окремими темами, надаючи структурований огляд відповідей відвідувачів. Однак LDA має свої обмеження в розумінні контексту та відтінків значень слів і фраз. Щоб вирішити цю проблему, ми інтегрували BERT - сучасну модель представлення мови, відому своєю здатністю вловлювати контекст слів у тексті, враховуючи слова, що стоять до і після речення. Підхід BERT, заснований на глибокому навчанні, дозволяє проводити більш контекстуальний аналіз відгуків, визначаючи "слова-маркери" - ключові терміни, які мають важливе значення або сенс, пов'язані з досвідом відвідувачів. Поєднання LDA та BERT забезпечує комплексний інструментарій для аналізу текстових відгуків від інтерактивних мистецьких інсталяцій, сценарії яких проаналізовано в [1, 2]. Результатом є глибше розуміння досвіду відвідувачів, їхніх емоційних реакцій та елементів інсталяцій, які резонують з ними найсильніше.

Дана стаття пропонує новий підхід до аналізу відгуків відвідувачів в інтерактивних мистецьких інсталяціях, використовуючи сильні сторони як LDA, так і BERT. Таким чином, маємо на меті надати художникам, кураторам і дослідникам цінну інформацію про вплив і сприйняття інтерактивних творів мистецтва, тим самим сприяючи розвитку дискурсу в галузі цифрового інтерактивного мистецтва.

1. State of the Art

Інтеграція семантичного аналізу даних в інтерактивне мистецтво зазнала значного прогресу за останні роки. Нові дослідження демонструють поєднання традиційного мистецтвознавчого аналізу з сучасними обчислювальними методами, зокрема, у

розумінні взаємодії відвідувачів та зворотного зв'язку в мистецьких інсталяціях.

Досягнення в галузі соціально-семантичного мережевого аналізу [3] заглиблюються в дуальність соціальних і семантичних мереж, підкреслюючи, як ці мережі впливають одна на одну. Їхнє розуміння соціально-семантичного мережевого аналізу підкреслює потенціал поєднання соціальних взаємодій із семантичним аналізом, що є актуальним для інтерактивного мистецтва, де зворотний зв'язок з відвідувачами охоплює як соціальний, так і індивідуальний досвід. Венскович і Норт [4] досліджують семантичну взаємодію у високорозмірних даних за допомогою підходу інтерактивного семантичного дослідження. Їхня методологія, зосереджена на кластеризації та проекції у двовимірному просторі, тісно пов'язана з процесом переробки складних відгуків відвідувачів у зрозумілі тематичні структури. Чжоу та ін. [5] пропонують ієрархічну модель взаємодії між модальностями для візуально-текстового аналізу настроїв, підкреслюючи важливість семантичної та емоційної взаємодії між різними модальностями. Це дослідження збігається з метою нашого дослідження - проаналізувати відгуки відвідувачів, які часто поєднують візуальні та текстові елементи. Штуббеманн, Дюрршнабель і Реффлінгхаус [6] обговорюють семантичний аналіз погляду у віртуальній і доповненій реальності, підкреслюючи важливість розуміння візуального сприйняття в інтерактивних умовах. Їхні висновки сприяють нашому розумінню того, як відвідувачі взаємодіють з інтерактивним мистецтвом, забезпечуючи основу для аналізу візуальної уваги та взаємодії. Сучасні дослідження в галузі семантичного аналізу даних, зокрема в контексті інтерактивного мистецтва, демонструють тенденцію до інтеграції різноманітних обчислювальних методів для розуміння складних людських взаємодій та досвіду. Методології варіюються від мережевого аналізу та інтерактивного дослідження даних до передового аналізу настроїв, кожна з яких дає цінну інформацію про відгуки відвідувачів та їхню залученість. Ці досягнення не лише збагачують наше розуміння досвіду відвідувачів в інтерактивному мистецтві, але й відкривають нові шляхи для міждисциплінарних досліджень у мистецтві, соціальних науках та комп'ютерному аналізі.

В даній статті розробляється та досліджується метод семантичного аналізу даних для визначення маркерних слів при обробленні результатів оцінки візиторів в інтерактивному мистецтві і забезпечення більш точного вибору. Ця задача розв’язуються в такій послідовності:

- на першому етапі відбувається попередня обробка даних. На цьому етапі з даних відфільтровуються зайві дані;

- на другому етапі відбувається обробка даних алгоритму LDA. На цьому етапі дані групуються на кластери;

- на третьому етапі відбувається пост обробка результату роботи алгоритму LDA алгоритмом BERT. На цьому етапі кластери з даними проходять додаткову обробку для виявлення більш точних маркерних слів для кожного кластера.

2. Попередня обробка даних

Попередня обробка даних є критично важливим етапом методології, який гарантує, що зібрані дані є чистими, структурованими та готовими до аналізу. Цей етап включає кілька кроків для перетворення необроблених текстових відгуків у формат, придатний для семантичного аналізу за допомогою латентного розподілу Діріхле (LDA) та двонаправленого кодувального представлення з трансформаторів (BERT). Ось детальний опис етапів попередньої обробки даних:

2.1. Нормалізація. Нормалізація в обробці тексту передбачає перетворення текстових даних у послідовний формат для полегшення точного аналізу [7]. У контексті підготовки даних для семантичного аналізу за допомогою C# на платформі .NET нормалізація зазвичай включає нормалізацію регістру, видалення діакритичних знаків (акцентів) і стандартизацію варіацій у тексті. Нормалізація регістру.

Мета. Забезпечити однорідність тексту, оскільки під час аналізу великі та малі літери одного і того ж слова повинні розглядатися як ідентичні.

Реалізація на C#. Використання методу ToLower() для перетворення всього тексту в нижній регістр.

2.2. Видалення діакритичних знаків (наголосів). *Мета*. У мовах з наголошеними символами часто буває корисно стандартизувати ці символи до їх базової форми, особливо коли семантичний аналіз не розрізняє наголошені та ненаголошені символи [8].

Реалізація на C#: Використання простору імен System.Text.Normalization та методу string.Normalize() для декомпозиції символів з діакритичними знаками, а потім видалення непідкреслених символів.

2.3. Стандартизація варіацій тексту. *Мета.* Обробка варіацій у тексті, які слід обробляти однаково, наприклад, синоніми, регіональні відмінності у написанні (наприклад, американська та британська англійська) або специфічна термінологія, що використовується як взаємозамінна [9].

Реалізація на C#. Створення словника варіацій та їхніх стандартизованих форм, потім заміна в тексті.

2.4. Токенізація. Токенізація – це важливий процес в аналізі тексту, під час якого текст розбивається на менші одиниці, як правило, слова або фрази. Ці одиниці, відомі як токени, є основою для подальшого аналізу, такого як синтаксичний, структурний або семантичний аналіз [10]. У C# токенізація може бути виконана за допомогою різних методів, залежно від складності тексту та вимог аналізу.

2.5. Розширена токенізація. *Мета.* Обробка складних текстових структур, таких як речення з розділовими знаками, скороченнями або іншими мовами з іншими правилами токенізації.

2.6. Видалення стоп-слів. Видалення загальних слів, які не роблять істотного внеску в зміст тексту, таких як "та", "і" і "або". Цей крок зменшує обсяг даних, що підлягають обробці, і допомагає зосередитися на словах, які мають більшу семантичну вагу [11].

2.7. Ідентифікація стоп-слів. Користувацький список стоп-слів залежно від конкретних потреб аналізу або тематики тексту може знадобитися створити власний список стоп-слів. Цей список може

включати специфічні для даної області терміни, які часто зустрічаються, але не є інформативними. Використання HashSet для зберігання стоп-слів є більш ефективним, ніж список або масив, оскільки забезпечує швидший час пошуку.

2.8. Лематизація. *Мета.* лематизація зводить слова до їхньої лематизованої форми, що передбачає складніший лінгвістичний аналіз для правильного приведення слова до його словникової форми. При цьому враховується частина мови, час та інші граматичні фактори [11]. *Реалізація на C#.* Ефективна реалізація лематизації вимагає всеосяжної лінгвістичної бази даних і складних алгоритмів, доступ до яких зазвичай можна отримати через бібліотеки НЛП. Однією з таких бібліотек є Stanford NLP, яку можна інтегрувати з додатками на C#.

2.9. Перетворення даних. Перетворення даних в обробці тексту - це важливий етап, на якому попередньо оброблений текст перетворюється в числовий формат, придатний для аналізу, особливо в моделях машинного навчання [11]. Цей процес включає в себе кілька ключових методів, кожен з яких призначений для обробки різних аспектів текстових даних.

2.10. Векторизація. *Мета.* Перетворення текстових даних у числові вектори, оскільки більшість алгоритмів потребують числових даних. *Метод.* Частота терміна, обернена до частоти документа (TF-IDF) – Подібний до BoW, але враховує частоту слова в усьому наборі даних, надаючи меншу вагу більш поширеним словам.

Завдяки ретельному виконанню цих етапів попередньої обробки дані ефективно перетворюються в чистий, структурований формат. Таке вдосконалення є важливим для подальших аналітичних етапів, щоб отримати точні та змістовні висновки з відгуків відвідувачів про інтерактивні мистецькі інсталяції.

3. Інтеграція LDA в обробку даних

Прихований розподіл Діріхле (Latent Dirichlet Allocation, LDA) - це популярна техніка моделювання тем, яка використовується в обробці природної мови (NLP) для виявлення абстрактних тем у колекції документів [12]. Впровадження LDA передбачає кілька ключових кроків, кожен з яких має важливе значення для вилучення значущих інсайтів з текстових даних. LDA ґрунтується на припущенні, що кожен документ є сумішшю різних тем і що кожна тема характеризується розподілом слів. Метою LDA є зворотне проектування цієї структури: враховуючи слова в документах, LDA намагається визначити набір тем, які, найімовірніше, згенерували б цю колекцію документів. Оцінка якості тем, згенерованих за допомогою LDA, є суб'єктивною, але дуже важливою. Для цього був вибраний наступний метод: Оцінка когерентності: Вимірює ступінь семантичної схожості між словами з високими показниками в темі. Вищі показники зв'язності зазвичай відповідають темам, які легше інтерпретувати.

Впровадження бібліотеки на основі Python, такої як Gensim для латентного розподілу Діріхле (LDA), у проекті на C# передбачає використання інтероперабельності між C# та Python. Цього можна досягти за допомогою таких інструментів, як Python.NET або IronPython. Python.NET є більш підходящим вибором для цього сценарію, оскільки він дозволяє C# взаємодіяти з Python та його бібліотеками напряму.

4. Інтеграція BERT в обробку даних

Двонаправлені кодові представлення з трансформаторів (BERT) – це метод, розроблений компанією Google для попереднього навчання NLP [13]. Інтеграція BERT в задачі обробки тексту включає кілька кроків, від початкового вибору моделі до кінцевої інтерпретації результатів. BERT призначений для розуміння контексту слова в реченні, дивлячись на слова, що стоять до і після нього. BERT вимагає вхідних даних у певному форматі.

Багато завдань можуть вирішуватися за допомогою попередньо навчених BERT-моделей безпосередньо або з додатковим налаштуванням. Такі бібліотеки, як Hugging Face's Transformers, забезпечують простий спосіб завантаження та використання цих моделей. Результати BERT можуть бути складними. Для задач вилучення ознак BERT надає вбудовування для кожного токена. Для задач класифікації вихідні дані з токена [CLS] можна подавати на додаткові шари, щоб отримати остаточну класифікацію.

5. Приклад практичного використання

Для ілюстрації зберемо набір даних, де кожен запис –опис емоцій відвідувачів картинної галереї.

Набір даних:

1. "Сьогодні я відчуваю себе надзвичайно щасливим!"
2. "Усередині мене глибокий смуток".
3. "Я киплю від гніву".
4. "Хвиля спокою накрила мене".
5. "Я здивований поворотом подій".

Кожне речення попередньо обробляється:

- приведення тексту до нижнього регістру;
- видалення розділових знаків та спецсимволів;
- розбиття тексту на слова;
- видалення стоп-слів (таких як "я", "є", "з");
- застосування лематизації.

Попередньо оброблені дані:

1. ["відчуваю", "абсолютно", "радісно", "сьогодні"];
2. ["глибоко", "смуток", "всередині"];
3. ["кипіння", "гнів", "несправедливість"];
4. ["хвиля", "спокій", "накритий"];
5. ["здивований", "поворот", "події"].

Реалізація LDA. Використовуючи LDA, ми визначаємо теми в цих описах емоцій. Ми встановили кількість тем 3. LDA може класифікувати дані наступним чином: Тема 1: ["радісний", "щасливий", "схвильований"] (позитивні емоції); Тема 2: ["сум", "горе", "печаль"] (негативні емоції); Тема 3: ["здивований", "шокований", "вражений"] ("емоції здивування").

Інтеграція BERT. Використовуючи BERT, ми генеруємо вставки для кожного попередньо обробленого речення, щоб вловити контекстні нюанси. Далі ми можемо використовувати ці вставки для таких завдань, як класифікація емоцій.

Встроювання BERT:

1. [0.85, -0.12, ...] (позитивні емоції);

2. [-0.76, 0.33, ...] (негативна емоція);
3. [-0,60, 0,29, ...] (негативна емоція);
4. [0,47, -0,15, ...] (позитивна емоція);
5. [0.22, 0.67, ...] (емоція здивування);

Поєднання результатів LDA та BERT. Після генерації тем LDA та вбудовувань BERT об'єднуємо ці результати, щоб отримати більш повне розуміння емоційного контексту кожного текстового запису. Для кожного речення ми пов'язуємо його з найбільш релевантною темою LDA і доповнюємо її контекстним розумінням, яке надають вбудовування BERT.

Комбінований аналіз:

1. Тема LDA: Позитивні емоції, BERT: [0,85, -0,12, ...] → Радісний, оптимістичний;
2. Тема ЛДА: Негативні емоції, BERT: [-0,76, 0,33, ...] → Сумний, меланхолійний;
3. Тема ЛДА: Негативні емоції, BERT: [-0,60, 0,29, ...] → Злий, розчарований;
4. Тема ЛДА: Позитивні емоції, BERT: [0.47, -0.15, ...] → Спокійний, миролюбний;
5. Тема ЛДА: Емоції здивування, BERT: [0.22, 0.67, ...] → Здивований, Заінтригований.

Об'єднані результати дають змогу глибше зрозуміти кожне речення. Теми LDA забезпечують широку категоризацію емоцій, тоді як вбудовування BERT пропонують нюансоване контекстно-залежне розуміння. Інтегруючи LDA і BERT, ми можемо ефективно аналізувати текстові дані, щоб виокремити як широкі тематичні елементи, так і тонкі контекстні нюанси емоцій людей. Такий підхід забезпечує більш глибоке і детальне розуміння, ніж будь-який з методів окремо, демонструючи силу поєднання різних технік NLP в аналізі текстів.

6. Обговорення

Основна методологія дослідження ґрунтується на синергетичному використанні LDA і BERT, двох сучасних технік NLP, кожна з яких робить свій унікальний внесок в аналіз текстового зворотного зв'язку. Модель LDA ефективно виокремлює широкі тематичні структури з відгуків, класифікуючи загальні настрої та теми, що переважають серед відвідувачів. На противагу цьому, роль BERT була ключовою в аналізі складних контекстуальних значень конкретних фраз і слів, що дозволило виявити слова-маркери, які інкапсулюють нюанси емоційних реакцій відвідувачів. Комбінований підхід забезпечив багатовимірне розуміння відгуків відвідувачів. Слова-маркери, визначені за допомогою цієї методології відобразили спектр реакцій від радості та здивування до роздумів і критики. Ці результати не лише підтверджують ефективність комплексного підходу, але й підкреслюють складність досвіду відвідувачів в умовах інтерактивного мистецтва. Для художників і кураторів ці висновки є досить важливими. Вони пропонують засновану на даних основу для розуміння залучення та реакції аудиторії. Це розуміння може вплинути на майбутні мистецькі творіння, дизайн виставок і навіть на кураторство інтерактивного досвіду, забезпечуючи глибший резонанс з аудиторією.

Поза межами інтерактивного мистецтва ця методологія має ширше застосування. Подібні підходи можна застосовувати в інших сферах, де розуміння суспільних настроїв і сприйняття має вирішальне значення, наприклад, в аналізі відгуків про продукт, аналізі настроїв у соціальних мережах та у тематичних дослідженнях у рамках якісних досліджень.

Дослідження визнає певні виклики. Обчислювальна інтенсивність BERT і необхідність ретельного налаштування параметрів у LDA є нетривіальними міркуваннями. Крім того, сфера дослідження була обмежена специфічним контекстом інтерактивного мистецтва, що може вплинути на узагальненість результатів. Майбутні дослідження можуть вивчити застосування цієї методології до різних наборів даних і контекстів для подальшого підтвердження її ефективності. Забігаючи наперед, можна сказати, що дослідження відкриває шляхи для включення більш складних методів NLP і вивчення семантичного аналізу, керованого ШІ. Удосконалення можуть включати аналіз зворотного зв'язку в реальному часі, крос-культурні порівняльні дослідження в інтерактивному мистецтві та інтеграцію мультимодального аналізу даних для включення візуального та слухового зворотного зв'язку поряд з текстовими даними.

Висновки

Досліджуючи сучасні методи аналізу тексту, ми заглибилися в синергетичну інтеграцію двох потужних технік NLP: прихованого розподілу Діріхле (LDA) та двонаправленого кодування за допомогою трансформаторів (BERT). Ця комбінація є значним кроком у галузі обробки природної мови, пропонуючи комплексний підхід до розуміння як широких тематичних структур, так і складних контекстуальних нюансів у текстових даних.

LDA довів свою ефективність у виявленні прихованої тематичної структури у великих текстових масивах, надаючи високорівневе уявлення про домінуючі теми. Його здатність розбивати величезні обсяги тексту на зрозумілі теми є безцінною для початкового розвідувального аналізу.

BERT робить наступний крок, аналізуючи текст на детальному рівні. Підхід, заснований на глибокому навчанні, до генерації вкладених слів фіксує тонкі контекстні значення слів на основі навколишнього тексту, що призводить до більш тонкого розуміння мови. Інтеграція LDA та BERT забезпечує можливість проведення дворівневого аналізу. У той час як LDA класифікує текст за ширшими темами, BERT забезпечує глибину, вловлюючи нюанси та складнощі мови, які можуть бути пропущені LDA. Цей комбінований підхід особливо ефективний у таких додатках, як аналіз настроїв, рекомендація контенту та емоційний аналіз.

Незважаючи на свою потужність, цей комплексний підхід не позбавлений проблем. Обчислювальна інтенсивність BERT, ретельне налаштування, необхідне для LDA, і потреба в надійних стратегіях попередньої обробки та валідації підкреслюють складність розширеного аналізу тексту.

Подальші дослідження і розробки можуть бути присвячені пошуку і аналізу метрик, які оцінюють якість визначення маркерних слів за допомогою

запропонованого та інших методів, а також розвиток і практичне впровадження напрацьованої методології для оброблення суджень-відповідей експертів, які надаються у вербальній формі, при розв'язанні важко формалізованих задач, зокрема, при оцінюванні безпеки [14].

СПИСОК ЛІТЕРАТУРИ

1. O. Golembovska, V. Kharchenko, I. Shostak, M. Danova, and O. Feoktystova. Assessing the Perception of Abstract Paintings with Elements of Augmented Reality, 11th IEEE DESSERT, Ukraine, 2020. DOI: 10.1109/DESSERT50317.2020.9125014.
2. O. Golembovska, V. Kharchenko, I. Shostak, M. Danova, O. Feoktystova, and V. Plietnov, Augmented Reality for the Abstract Paintings: Application Scenarios, Semantic Similarity Analysis and Case Study, *2019 10th IEEE Int. Conf. on IDAACS.: Technology and Applications*, Metz, France, 2019, pp. 1007-1011, DOI: 10.1109/IDAACS.2019.8924411.
3. N. Basov, R. Breiger, I. Hellsten. Socio-semantic and other dualities. Poetics. 2020. p.101433, DOI: 10.1016/j.poetic.2020.101433.
4. Pollux: Interactive Cluster-First Projections of High-Dimensional Data [Текст] / John E. Wenskovitch, C. North // 2019 IEEE Visualization in Data Science (VDS) - 2019. – pp.38-47, DOI: 10.1109/VDS48975.2019.8973381.
5. Visual-Textual Sentiment Analysis Enhanced by Hierarchical Cross-Modality Interaction [Текст] / Tao Zhou, Jiuxin Cao, Xueling Zhu, Bo Liu, Shancang Li // IEEE Systems Journal - 2021. – pp.4303-4314, DOI: 10.1109/jsyst.2020.3026879.
6. Neural Networks for Semantic Gaze Analysis in XR Settings / Lena Stubbemann, Dominik Dürrschnabel, R. Refflinghaus // ACM Symposium on Eye Tracking Research and Applications - 2021, DOI: 10.1145/3448017.3457380.
7. Multilingual Sequence Labeling Approach to solve Lexical Normalization / Divesh R. Kubal, Apurva Nagvenkar // 2021 The 7th Workshop on Noisy User-generated Text (W-NUT) - 2021. – p.457-464, DOI: 10.18653/v1/2021.wnut-1.51.
8. Proposed Natural Language Processing Preprocessing Procedures for Enhancing Arabic Text Summarization / Reda Elbarougy, G. M. Behery, Akram el Khatib // 2019. – p.39-57, DOI: 10.1007/978-3-030-34614-0_3.
9. The influence of preprocessing on text classification using a bag-of-words representation / Yaakov Hacohen-Kerner, Daniel Miller, Yair Yigal // PLoS ONE - 2020. – DOI: 10.1371/journal.pone.0232525.
10. Italian Text Categorization with Lemmatization and Support Vector Machines / F. Camastra, Gennaro Razi // 2020. – p.47-54, DOI: 10.1007/978-981-13-8950-4_5.
11. From Words to Numbers: Getting Started with Text Analysis for Applied Social Scientists [Текст] / Hyun Woo Kim, Hyejung Chang // BCRP (Business Communication Research and Practice) - 2020. – p.122-129, DOI: 10.22682/BCRP.2020.3.2.122.
12. A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura / S. Zhou, P. Kan, Qunying Huang, J. Silbernagel // Journal of Information Science - 2021. DOI: 10.1177/01655515211007724.
13. Neural Topic Models for Short Text Using Pretrained Word Embeddings and Its Application To Real Data / R. Murakami, B. Chakraborty. 2021 IEEE 4th Int Conf on Knowledge Innovation and Invention (ICKII) - 2021. – p.146-150, DOI: 10.1109/ICKII51822.2021.9574752.
14. Babeshko, I.; Leontiiev, K.; Kharchenko, V.; Kovalenko, A.; Brezhniev, E. Application of Assumption Modes and Effects Analysis to XMECA. In Theory and Engineering of Dependable Computer Systems and Networks; DepCoS-RELCOMEX 2021. Advances in Intelligent Systems and Computing; Springer: Cham, Switzerland, 2021; Volume 1389. DOI: 10.1007/978-3-030-76773-0_1.

Received (Надійшла) 21.12.2023
Accepted for publication (Прийнята до друку) 07.02.2024

Method of semantic data analysis for determining marker words in the processing of visitors' evaluation results in interactive art

V. Narozhnyi, V. Kharchenko

Abstract. The subject of the study is in-depth semantic data analysis based on the integration of the methodologies of latent Dirichlet distribution (LDA) and bidirectional encoding representation from transformers (BERT). This research focuses on processing textual data, in particular, visitors' evaluations of interactive art, to identify marker words that highlight key emotional and thematic elements. The goal is to deepen the understanding of visitors' experiences and perceptions of interactive art installations by identifying significant marker words using a combined LDA and BERT approach. This combination aims to capture both general thematic content and the nuanced context of feedback. Objectives: collection and preprocessing of textual data - visitor ratings, consisting of tokenization, normalization and lemmatization steps with the implementation of LDA to extract common themes from the collected data, providing insights into the main themes present in visitor feedback; integration of BERT to analyze contextual nuances and extract deeper meanings from individual words in the feedback; combining the results of LDA and BERT to create a comprehensive understanding of the textual data, focusing on identifying the most significant marker words. The following results were achieved successful extraction of key themes from visitors' ratings using LDA, which allowed us to identify broad thematic categories present in the reviews; a deep learning approach BERT was proposed, which provided nuanced contextual embeddings, emphasizing specific emotions and sentiments expressed by visitors; the results of LDA and BERT were integrated, which provided a rich set of marker words that effectively reflect the essence of the experience and perception of visitors to interactive art; the accuracy and depth of analysis in identifying key emotional and thematic elements was improved, as evidenced by the consistency and relevance of marker words in relation to visitors' ratings. Conclusions: The integration of LDA and BERT for semantic data analysis in interactive art contexts demonstrates a powerful approach for understanding complex visitor feedback. This method provides a two-level analysis, where LDA offers insights into general themes and BERT contributes to detailed contextual understanding. The study successfully identifies specific marker words that effectively capture the essence of visitors' impressions and ratings. This methodology can be useful for artists, curators, and researchers in measuring public reception and improving interactive art experiences. The adaptability of the methodology creates real prospects for its application in other areas that require a detailed semantic analysis of textual feedback.

Keywords: semantic data analysis, natural language processing, latent Dirichlet distribution, bidirectional coded representations from transformers, interactive art, emotional response analysis.

УДК 004. 891 + 330.46 **doi: 10.26906/SUNZ.2024.1.146**

Л. О. Нікітіна[1], Н. В. Дженюк[1], Л. В. Борисова[2]

[1] Національний технічний університет «Харківський політехнічний інститут», Харків, Україна
[2] Національний університет цивільного захисту України, Харків

ЕКСПЕРТНА СИСТЕМА ДЛЯ ОЦІНКИ РИЗИКІВ ХМАРНИХ СЕРВІСІВ

Анотація. Реалії сучасності вимагають від суспільства впровадження цифрових технологій, які набувають все більшої складності та інтелектуальності. Цифровізація (digitalization) стає невід'ємним компонентом усіх сфер діяльності людини. Тенденцією розвитку та економічного зростання фірм і організацій стають хмарні технології, які дозволяють організаціям мати гнучкі витрати в IT-секторі і регулювати їх шляхом купівлі доступу до ресурсів та сервісів у провайдерів замість купівлі самих ресурсів та сервісів. Для організацій, які приймають рішення, пов'язані з використанням хмарних сервісів, виникають труднощі з оцінкою та оптимальним вибором сервісів та провайдерів, оскільки для цього поки що не існує загальноприйнятих вказівок або процедур. З іншого боку, перед провайдерами постає проблема забезпечення належної якості хмарних сервісів, що надаються користувачам. Як провайдерам, так і користувачам необхідно мати інструменти, які дають змогу визначити та оцінити можливі ризики хмарних сервісів. Одним з таких інструментів може бути експертна система з оцінки хмарних сервісів, концепція якої розглядається у даній статті.

Ключові слова: хмарні обчислення; хмарні послуги; експертна система; система нечіткого логічного висновку; база знань; оцінка ризиків; вразливість.

Вступ

Хмарні технології в наш час набувають все більшої популярності. В області інформаційних технологій термін "хмара" використовується для позначення хмарних обчислень або хмарних сервісів. У «хмарі» поєднуються комунікації, ресурси апаратного та програмного забезпечення, сховища даних. Доступ до ресурсів забезпечується через мережу Інтернет. Користувачі «хмари» не повинні прямо володіти фізичним обладнанням або управляти ним, вони можуть віддалено використовувати ресурси за потребою та платити лише за фактичне використання. Крім того, коли не вистачає потужності власних ресурсів, користувачі можуть розгортати свої застосунки у «хмарі», і використовувати їх у зручний спосіб.

Будемо використовувати такі означення [1, 2]:

– хмарні технології – концепція надання послуг зі зберігання та обробки даних, згідно з якою обчислювальні ресурси надаються користувачеві через Інтернет як онлайн сервіси;

– хмарний сервіс – послуга з надання хмарних ресурсів за допомогою технологій «хмарних обчислень»;

– хмарні обчислення – використання обчислювальних служб (серверів, сховищ, баз даних, комунікаційних мереж, програмного забезпечення, аналітики та інтелектуального аналізу) через мережу Інтернет в режимі «на вимогу» згідно з угодою з провайдером, який надає ці послуги;

– сервіси – служби, які забезпечують хмарні обчислення; вони дозволяють прискорити впровадження інновацій, підвищити гнучкість використання ресурсів, отримати економію завдяки високій масштабованості; користувач зазвичай платить лише за хмарні сервіси у міру зміни потреб бізнесу;

– постачальник хмарних послуг (CSP, сервіс-провайдер) – це IT-компанія, яка надає масштабовані обчислювальні ресурси на вимогу, наприклад обчислювальну потужність, сховище даних або програми через Інтернет.

Хмара може бути організована як сукупність великої кількості фізичних хостів, які об'єднані телекомунікаційною мережею. Прикладами постачальників хмарних послуг є Microsoft Azure (120 зон доступу, 62 регіони), Google Cloud Platform (GCP, 118 зон, 39 регіонів), Amazon Web Services (AWS, 102 зони, 32 регіони), Alibaba Cloud (89 зон, 30 регіонів), Oracle Cloud (46 зон, 38 регіонів), IBM Cloud (30 зон, 10 регіонів) та ін. [2–9]. Комерційні хмари можуть працювати на мільйонах фізичних хостів. Кожен із цих хостів може розміщувати багато віртуальних машин (VM), за допомогою яких їх можна викликати або видаляти динамічно. Крім того, хмарні гіпервізори використовуються для керування наданням ресурсів, що передбачає відображення та планування створених віртуальних машин, що знаходяться на фізичних серверах хмари. Ці технічні та економічні переваги хмарних обчислень на вимогу зробили можливим переміщення традиційних корпоративних обчислень до хмар. Хмарні технології мають ряд переваг: зручність, доступність, економія коштів за рахунок зниження вартості інфраструктури, масштабованість, певний рівень безпеки даних. Однак, окрім явних переваг, вони мають і недоліки: обмежена пропускна здатність, залежність від провайдера, проблеми з приватністю, привабливість для зловмисників. Зазначені недоліки роблять актуальною проблему визначення та оцінки ризиків у організації та використанні хмарних обчислень. Такі ризики мають враховувати як провайдери хмарних сервісів, так і користувачі (організації і приватні особи) при виборі сервісів та провайдерів.

1 Хмарні обчислення

Хмарні обчислення будуються на основі клієнт-серверної моделі, метою якої є підвищення доступності обчислювальних ресурсів. Національний інститут стандартів і технологій (National Institute of Standards and Technology, NIST) визначає, що хмарні обчислення мають ряд характеристик і будуються на основі моделі сервісів і моделі розгортання [10–12]. Головними характеристиками хмарних обчислень є:

1) об'єднання ресурсів – постачальник хмарних послуг може розподіляти ресурси між кількома клієнтами, кожен з яких використовує свій набір послуг;

2) самообслуговування на вимогу – клієнт за необхідності без взаємодій з персоналом постачальника послуг може задіяти обчислювальні можливості (серверний час, мережеве сховище даних та ін.), безперервно відстежувати та контролювати обчислювальні можливості відповідно до своїх потреб;

3) легкість обслуговування – оновлення та оптимізація ресурсів хмари (серверів) відбувається за мінімальний або навіть нульовий час;

4) масштабованість і гнучкість – можливість швидко та ефективно балансувати поточні навантаження, які потребують великої кількості серверів;

5) економність – зменшення витрат на організацію та використання простору для зберігання даних, бо найчастіше він виділяється безкоштовно;

6) служби вимірюваності послуг та звітності – дозволяють провайдерам і клієнтам відстежувати, які послуги та з якою метою використовувалися, та формувати відповідні звіти; це допомагає контролювати виставлення рахунків і забезпечувати оптимальне використання ресурсів;

7) безпека – хмарні служби створюють резервні копії даних, щоб запобігти будь-якій втраті даних;

8) автоматизація – здатність автоматично встановлювати, налаштовувати та підтримувати хмарні сервіси відома як автоматизація в хмарних обчисленнях; це вимагає встановлення та розгортання віртуальних машин, серверів і великих сховищ;

9) стійкість – здатність сервісу швидко відновлюватися після будь-яких збоїв, час перезапуску та відновлення серверів, баз даних і мережевих систем після будь-яких втрат або пошкоджень;

10) доступність – доступ до хмарних сервісів можна отримати віддалено, без географічних обмежень або обмежень на використання хмарних ресурсів;

11) широкий доступ до мережі – клієнти можуть отримати доступ до хмарних даних або перенести дані в хмару з будь-якого місця за допомогою пристрою та підключення до Інтернету.

Хмарні сервіси управляють доступом до ресурсів хмари відповідно до вимог клієнта.

Хмарні обчислення пропонують такі три види сервісів:

1) програмне забезпечення як сервіс (SaaS, сервіси хмарних додатків) – здебільшого додатки SaaS запускаються безпосередньо через веб-браузер, і користувачеві не потрібно завантажувати та встановлювати ці програми; за допомогою SaaS користувач може отримати доступ до програмного забезпечення через Інтернет без потреби в будь-якій платформі; приклади: Google Apps, Salesforce Dropbox, Slack, Hubspot, Cisco WebEx;

2) платформа як сервіс (PaaS) – надає платформу для створення програмного забезпечення; приклади: Windows Azure, Force.com, Magento Commerce Cloud, OpenShift;

3) інфраструктура як послуга (IaaS) – відповідає за керування даними, які використовуються програмними додатками, проміжним програмним забезпеченням (middleware) і середовищами виконання; приклади: Amazon Web Services (AWS) EC2, Google Compute Engine (GCE), Cisco Metapod.

Під терміном "модель розгортання хмарного обчислення" розуміють архітектуру розгортання віртуального обчислювального середовища, що змінюється залежно від обсягу даних, які ви треба зберегти, і того, хто має доступ до інфраструктури. Модель розгортання хмари визначає конкретний тип хмарного середовища, на основі характеристик типу власності, масштабу, типу доступу, характеру, призначення хмари, розташування серверів і функцій керування ними.

При виборі провайдера та сервісів користувачеві необхідно зрозуміти, яка модель найкраще йому підходить в конкретних умовах для вирішення конкретних задач.

Існують такі основні моделі розгортання хмарних обчислень:

– публічна хмара (Public Cloud) – є доступною для всіх, і будь-яка організація може мати доступ до систем і сервісів;

– приватна хмара (Private Cloud) – послуги побудовані відповідно до принципів хмарних обчислень, але доступні лише в приватній мережі;

– хмара спільноти (Community or Partner Cloud) – хмарні послуги провайдер пропонує обмеженій і чітко визначеній кількості сторін.

Крім того, існують моделі гібридної хмари (Hybrid Cloud, об’єднання приватних та публічних хмар), багатопровайдерної хмари (Multi-Cloud, комбінація приватних хмар, публічних хмар або приватних і публічних хмар).

Кожну з основних моделей можна оцінити за чотирибальною шкалою за такими параметрами як "товарність" (commodity), вартість (cost), відповідальність (liability) і гарантованість (assurance) у порівнянні з "не-хмарою" (табл. 1, [13]).

На даний момент не існує загального підходу для вибору моделі розгортання хмари. Вибір моделі необхідно робити відповідно до поточних вимог. При виборі найкращої моделі розгортання можна враховувати, крім зазначених вище, такі фактори як масштабованість (Scalability), легкість використання (Easy to use), конфіденційність (Privacy), відповідність (Compliance) та ін.

Таблиця 1 – **Параметри для оцінки моделей**

	"товарність" (commodity)	Вартість (cost)	відповідальність (liability)	гарантованість (assurance)
Public Cloud	4	1	1	1
Private Cloud	3	2	2	2
Partner Cloud	2	3	3	3
Non-Cloud	1	4	4	4

Ризики та вигоди, пов’язані з кожною моделлю хмарних обчислень, відрізняються і користувачеві це треба усвідомлювати і враховувати при виборі провайдера та сервісів.

2 Загрози, вразливості та ризики у хмарних обчисленнях

Безпека використання хмарних обчислень полягає у забезпеченні доступності, цілісності, конфіденційності та підтримці інформаційних ресурсів інфраструктури. Будь-яка фірма, яка використовує хмару, прагне захистити свої активи, найціннішим з яких є інформація. Хмарні сервіси надають можливість отримати швидкий та зручний доступ до інформаційних ресурсів та сервісів, але, водночас, такі ресурси і сервіси можуть бути вразливими до різноманітних небезпек і загроз. Вразливості хмари стають причинами широкого спектру ризиків, які впливають на активи як користувачів, так і провайдерів хмарних сервісів (рис. 1).

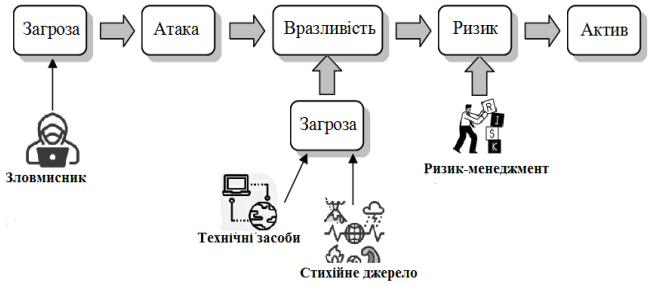


Рис. 1. Зв'язок між загрозами, вразливостями, ризиками

Загроза (threat) у контексті хмарних обчислень – це атака на хмарні ресурси, яка намагається порушити їхню роботу та доступ до них. Загрози мають широкий спектр дії – від втрати та витоку даних, випадкового розкриття облікових даних до складних кібератак. Загрози спрямовані на найбільш слабкі ланки системи захисту хмари – на вразливості і призводять до порушення інформаційної безпеки, режиму функціонування і доступності конкретних компонентів системи.

Хмарні вразливості (vulnerabilities) – це слабкі місця або прогалини в середовищі хмарних обчислень, якими зловмисники можуть скористатися, щоб отримати несанкціонований доступ, викрасти дані або порушити роботу сервісів.

Ризики у хмарних обчисленнях є результатом вразливостей хмарних ресурсів та сервісів під дією атак. Відповідальність за пом'якшення ризиків розподіляється між провайдером і хмарним споживачем.

Ризик – це можливість втрати, пошкодження або знищення активів або даних особи або організації через загрозу подій або дій. Ризики виникають як результат вразливостей.

Організація може бути вразливою до різноманітних загроз, які можуть впливати на ефективність роботи та дотримання нормативних вимог. Щоб запобігти впливу ризиків простого усвідомлення недостатньо. Необхідно використовувати управління ризиками у хмарі (CRM) для їхньої мінімізації, а в деяких випадках і усунення [13-18].

Найважливіші класи ризиків, пов'язаних із хмарою (не у порядку критичності) [13]:

1) втрата керування: при використанні хмарних інфраструктур клієнт обов'язково передає контроль постачальнику хмарних технологій (СР) щодо низки питань, які можуть вплинути на безпеку; при цьому згідно з угодою про рівень обслуговування провайдер хмарних послуг може бути не зобов'язаним надавати такі послуги, а це є прогалиною в безпеці;

2) блокування: відсутність доступних інструментів, процедур або стандартних форматів даних чи інтерфейсів послуг, які могли б гарантувати переносимість даних, програм і послуг; це може ускладнити або заблокувати для клієнта перехід від одного постачальника до іншого або перенесення даних і послуг назад у власне IT-середовище;

3) помилки ізоляції: ця категорія ризику охоплює збій механізмів, що розділяють сховище, пам'ять, маршрутизацію та навіть репутацію між різними орендарями (наприклад, так звані атаки з переходом на гостьову систему);

4) ризики відповідності: інвестиції в отримання сертифікації (наприклад, галузевого стандарту чи нормативних вимог) можуть бути піддані ризику через міграцію до хмари, якщо СР не може надати докази власної відповідності відповідним вимогам або якщо СР не дозволяє аудит клієнтом хмари (СС);

5) компроміс інтерфейсу керування: інтерфейси керування клієнтами постачальника загальнодоступної хмари доступні через Інтернет і забезпечують доступ до більших наборів ресурсів (ніж у традиційних хостинг-провайдерів), тому становлять підвищений ризик, особливо в поєднанні з віддаленим доступом і вразливістю веб-браузера;

6) захист даних: у деяких випадках замовнику хмари (у ролі контролера даних) може бути важко ефективно перевірити практику обробки даних постачальником хмари і переконатися, що дані обробляються належним чином, особливо у випадках багаторазової передачі даних, наприклад, між об'єднаними хмарами;

7) небезпечне або неповне видалення даних: запит на видалення хмарного ресурсу не завжди виконується як справжнє видалення даних, додаткові копії даних зберігаються, але вони недоступні, або тому, що диск, який потрібно знищити, також зберігає дані з інших клієнтів;

8) зловмисний інсайдер: хмарні архітектури вимагають певних ролей, які є надзвичайно ризикованими, наприклад, ролі системних адміністраторів СР і постачальників послуг керованої безпеки;

Стандарт ISO/IEC 27001:2022 регламентує стратегію інформаційної безпеки, орієнтовану на захист конфіденційності, забезпечення автентичності і доступності даних [28]. Аналіз та інтерпретація ризику виконуються за допомогою оцінки ризику. Цей процес базується на виявленні та оцінці вразливостей, які існують в організації [14]. В стандарті ISO 31000:2018 акцент концепції ризику робиться не тільки на визначенні його ймовірності та наслідків, а й на процесі управління ризиками. Управління ризиками в хмарі – це процес оцінки, захисту та керування ризиками, пов'язаними із хмарними обчисленнями. Управління ризиками визначає, які проблеми мають пріоритет і як реагувати на можливі ризики. Процес управління ризиками орієнтований

на врахування потенційних небезпек, які стосуються як провайдерів, так і користувачів [29].

В деяких випадках клієнту хмари доцільно і можливо передавати ризик постачальнику хмари; однак не всі ризики можна передати: якщо ризик призводить до краху бізнесу, серйозної шкоди репутації або юридичних наслідків, будь-якій іншій стороні важко або неможливо компенсувати цю шкоду.

3 Експертна система для оцінки ризиків хмарних сервісів

Для організацій, які приймають рішення, пов'язані з використанням хмарних сервісів, виникають труднощі з оцінкою та оптимальним вибором сервісів та провайдерів, оскільки для цього поки що не існує загальноприйнятих вказівок або процедур. З іншого боку, перед провайдерами постає проблема забезпечення належної якості хмарних сервісів, що надаються користувачам. Як провайдерам, так і користувачам необхідно мати інструменти, які дають змогу визначити та оцінити можливі ризики хмарних сервісів. Одним з таких інструментів може бути експертна система (ЕС) для оцінки хмарних сервісів, концепція якої розглядається у даній статті.

Експертна система дозволяє:

– сформувати базу даних:

o реєстр можливих вразливостей хмарних сервісів;

o реєстр ризиків;

o реєстр активів користувача;

o таблиці оцінок рівнів ризиків та їхнього впливу на активи;

o таблиці ймовірностей ризиків;

o сценарії реагування на ризики;

– сформувати базу знань на основі продукційної моделі;

– виконати оцінку та аналіз ризиків;

– провайдеру – отримати рекомендації з формування реакції на ризики;

– користувачеві – визначити та порівняти вплив ризиків, притаманних різним провайдерам, на активи користувача;

– зберігати у базі даних звіти, сформовані у ході сеансів роботи системи.

Така система може бути побудована за архітектурою експертних систем (рис. 2).

Користувачами ЕС є привілейовані користувачі з боку провайдера – експерт і ризик-менеджер та кінцевий користувач хмарних сервісів. Привілейовані користувачі мають доступ до формування бази даних та знань через підсистему управління знаннями та до підсистеми управління ризиками. Кінцевий користувач може формувати вхідні дані для виконання оцінки ризиків та їхнього впливу на важливі для нього активи.

Експертна система, запропонована в цьому документі, містить інтерфейси, з якими взаємодіють користувачі, машину виведення, яка виконує обґрунтування знань/даних, базу даних і базу знань, яка зберігає загальні та абстрактні знання про оцінку комерційних хмарних сервісів. База знань формується на основі знань експертів та знань про хмарні обчислення, опубліковані в джерелах інформації.

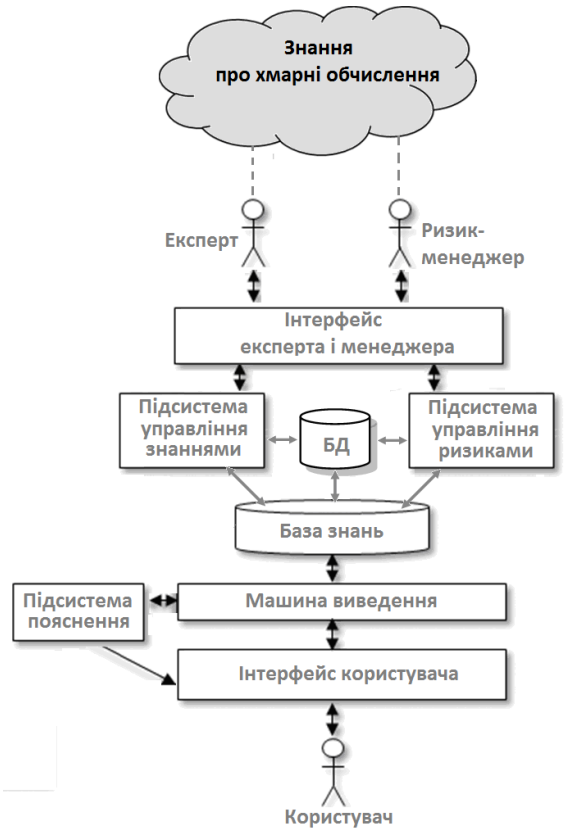


Рисунок 2 – Архітектура експертної системи

Наразі ми зробили акцент на формуванні бази знань та роботі машини виведення для оцінки ризиків. Програмна реалізація системи у даному документі не розглядається.

4 База знань та машина виведення

Вивчення джерел інформації [10-27] дало підставу зробити висновок про доцільність побудови бази знань на основі продукційної моделі.

Правило продукції – це вираз виду:

$$(i): Q; P; A; \Rightarrow B; S, F, N, \qquad (1)$$

де (i) – унікальний ідентифікатор продукції; Q – сфера застосування продукції; P – умова застосовності ядра продукції; $A \Rightarrow B$ – ядро продукції, в якому A – умова ядра, B – висновок ядра; "$\Rightarrow$" – знак логічної секвенції (наслідку); S – метод або спосіб визначення кількісного значення ступеню істинності висновку ядра; F – коефіцієнт визначеності або впевненості продукції; N – післяумова продукції.

База знань складається з множини правил виду (1). Кожне правило являє собою незалежну одиницю знань. Передумови можуть розглядатися як модель (образ), а наслідок – як висновки або дії, які необхідно виконати. Для отримання експертизи хмарних сервісів користувач ЕС налаштовує базу знань, вводить вхідні дані (наявні або гіпотетичні значення ступеню вразливостей хмари, важливість своїх активів, ризики, що перебувають у сфері його інтересів та ін.) та запускає машину виведення.

На основі рекомендацій ЕС користувач може прийняти рішення про перехід до хмарного середо-

вища певного провайдера або порівняти ризики різних провайдерів (рис. 3).

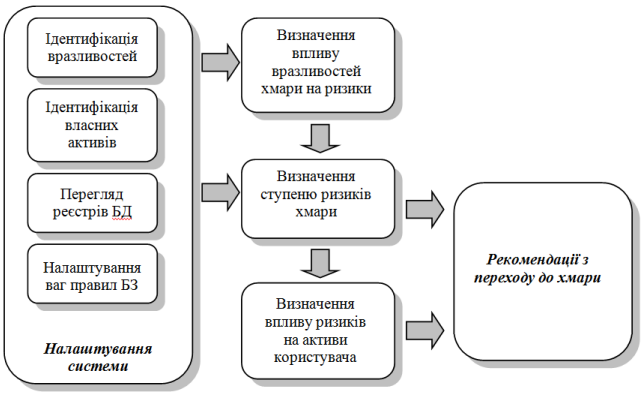


Рис. 3. Рекомендації користувачеві хмари

У циклі виведення виконуються такі операції:

1) співставлення – зразок правила співставляється з фактами, наявними у базі фактів;

2) вибір – якщо знайдено декілька підходящих правил, то вони створюють конфліктний набір; з конфліктного набору вибирається одне правило, яке найбільше підходить за заданим критерієм – тобто виконується рішення конфлікту;

3) спрацьовування – якщо співставлення антецедента правила з фактами робочої пам'яті виконано успішно, то правило спрацьовує;

4) дія – до робочої пам'яті додається новий істинний факт, що є консеквентом правила, яке спрацювало.

Результатом є визначення впливу вразливостей на ризики і ризиків – на активи.

Через інтерфейс користувачеві ЕС доступні функції виконання запитів до БД та БЗ: перегляд наявних реєстрів вразливостей, ризиків, активів, перегляд наявних правил та їхньої ваги, виконання різноманітних вибірок. Для визначення готовності до хмарного середовища користувачеві потрібно зібрати необхідну інформацію про:

– провайдера хмарних сервісів;
– сторонніх постачальників;
– поточні рішення та конфігурацію безпеки.

Для повного контрольного списку оцінки ризиків хмари користувачеві необхідно виконати:

1) визначення всіх активів, які зберігатимуться у хмарному середовищі – дані клієнтів, фінансових записів, облікові дані співробітників, відомості про комерційну діяльність;

2) класифікацію своїх даних відповідно до їх чутливості; це допоможе визначити активи, які піддаються найбільшому ризику та потребують кращого захисту;

3) визначення потенційних загроз; тестування хмарних загроз і проникнення найкраще доручити експертам, які знайомі з векторами атак і мають інструменти, необхідні для моделювання атак;

4) оцінювання ризиків, пов'язаних з кожною загрозою та впливу на активи.

Привілейовані користувачі (експерт та менеджера з ризиків) крім функцій, доступних функції користувачеві, можуть отримати рекомендації з управління ризиками (рис. 4).

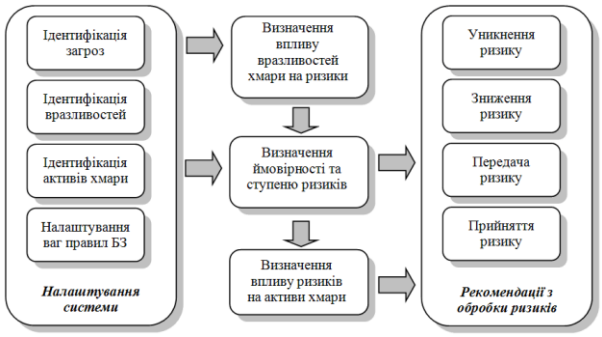


Рис. 4. Рекомендації менеджеру для обробки ризиків

Вважаємо доцільним створення нечіткої бази знань. Одиницями знань у ній будуть нечіткі правила. Приклади таких правил наведені у табл. 2:

Таблиця 2 – **Приклади правил**

Ідентифікатор правила		Антецедент		Консеквент
001	IF	V46=H, V47=M, V31=L	THEN	R05=M
…		…		…
…	IF	R05=H, R09=H, R15=M, R18=M, R32=L	THEN	A10=L

У наведеному прикладі використано позначення: V_n – ідентифікатор вразливості згідно з реєстром вразливостей; R_m – ідентифікатор ризику згідно з реєстром ризиків; A_k – ідентифікатор активу згідно з реєстром активів; H – високий рівень впливу; M – середній рівень впливу; L – низький рівень впливу.

Машина виведення може бути організована як система нечіткого виведення за алгоритмами Мамдані, Сугено та ін. з використанням різних способів дефаззіфікації результатів.

Іншим варіантом може бути багатоступенева нейро-нечітка система виведення. Така система має бути попередньо навчена на відповідних зразках для визначення рівнів ризиків та їхнього впливу на активи.

Висновки

Бурхливий розвиток хмарних обчислень викликав появу комерційних постачальників хмарних сервісів. Умови та спектр пропозицій, характеристики пропонованих сервісів можуть мати суттєві відмінності та особливості. З цієї причини є важливим перед використанням хмарних сервісів певного провайдера проводити оцінювання потенціальних ризиків і їхніх впливів на наявні активи. Крім того, вибираючи з кількох потенційних провайдерів хмарних сервісів, можна порівнювати їх між собою.

Сфера створення та надання хмарних послуг стрімко змінюється, певні аспекти безпеки стають неконтрольованими клієнтами, тому зростають ризики використання сервісів.

Оцінка комерційних хмарних послуг неминуче стає більш складною, ніж оцінка традиційних обчислювальних систем. Для полегшення роботи з оцінювання ризиків у контексті хмарних обчислень та

впливу їх на активи користувача ми запропонували створити експертну систему на основі накопичення та застосування наявних експертних знань у галузі хмарних обчислень. Запропонована експертна система може бути використана як інструмент з надання рекомендацій як провайдерам хмарних послуг в результаті оцінки ризиків, так і користувачам при виборі провайдера.

СПИСОК ЛІТЕРАТУРИ

1. Peter Mell Timothy Grance. The NIST Definition of Cloud Computing. Recommendations of the. NIST Special Publication 800-145. September 2011. https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf
2. Cloud computing. IT Enterprise. https://www.it.ua/knowledge-base/technology-innovation/cloud-solutions
3. What is cloud computing? https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cloud-computing
4. What is cloud computing? https://www.ibm.com/topics/cloud-computing
5. What is Cloud Computing? https://cloud.google.com/learn/what-is-cloud-computing
6. Top 5 Cloud Services Providers 2023. https://savemyleads.com/blog/useful/top-5-cloud-services-providers-2023
7. Що таке хмарні технології? Переваги та недоліки. https://edin.ua/shho-take-xmarni-texnologi%D1%97-i-navishho-voni-potribni/
8. The top 10 public cloud providers in 2023. https://www.revolgy.com/insights/blog/the-top-10-public-cloud-providers-2023
9. Top 10 Cloud Service Providers Globally in 2023. https://dgtlinfra.com/top-cloud-service-providers/
10. Nayan Ruparelia. Cloud computing. Cambridge, MA : The MIT Press, 2016 – 278 p. https://s3.amazonaws.com/arena-attachments/911381/0ea8a9793158a95d9b91911e49240a43.pdf
11. T.B. Rehman. Cloud Computing Basics. MERCURY LEARNING AND INFORMATION. Mercury Learning and Information LLC, 2019 – 198 p. https://terrorgum.com/tfox/books/cloudcomputingbasics_aselfteachingintroduction.pdf
12. Cloud Computing. https://www.javatpoint.com/cloud-computing
13. ENISA. Cloud computing: benifits, risks and recommendation for information security. Nov 09. https://www.enisa.europa.eu/publications/cloud-computing-risk-assessment
14. Fotis Kitsios, Elpiniki Chatzidimitriou, Maria Kamariotou. The ISO/IEC 27001 Information Security Management Standard: How to Extract Value from Data in the IT Sector. https://www.researchgate.net/publication/369606652_The_ISOIEC_27001_Information_Security_Management_Standard_How_to_Extract_Value_from_Data_in_the_IT_Sector
15. INTERNATIONAL STANDARD. ISO/IEC 27017. Information technology — Security techniques — Code of practice for information security controls based on ISO/IEC 27002 for cloud services.
16. Risk Management in Cloud Computing. https://www.scrut.io/post/risk-management-in-cloud-computing
17. Pedro Costa, João Paulo Santos, Miguel Mira da Silva. Evaluation Criteria for Cloud Services. https://www.researchgate.net/publication/261436007_Evaluation_Criteria_for_Cloud_Services
18. Timothy Morrow. 12 Risks, Threats, & Vulnerabilities in Moving to the Cloud. https://insights.sei.cmu.edu/blog/12-risks-threats-vulnerabilities-in-moving-to-the-cloud/
19. Shannon Jackson-Barnes. Cloud Computing: Common Vulnerabilities and How to Overcome Them. https://www.orientsoftware.com/blog/vulnerability-in-cloud-computing/
20. Nivedita James Palatty Cloud Vulnerability Management: The Detailed Guide. https://www.getastra.com/blog/security-audit/cloud-vulnerability-management/
21. What Is Cloud Vulnerability Assessment And How To Implement It? https://discovercloud.io/what-is-cloud-vulnerability-assessment-and-how-to-implement-it/
22. Saumick Basu. 5 Top Cloud Vulnerability Scanners for AWS, Google Cloud, and Azure. https://www.getastra.com/blog/security-audit/cloud-vulnerability-scanner/?nowprocket=1
23. A Comprehensive Guide to Cloud Vulnerability Management. https://www.clouddefense.ai/guide-to-cloud-vulnerability-management/
24. Cloud Vulnerability Management Best Practices for 2024. https://www.sentra.io/learn/cloud-vulnerability-management
25. Martin Zboril. RISK ASSESSMENT METHOD OF CLOUD ENVIRONMENT. Computing and Informatics, Vol. 41, 2022, 1186–1206, doi: 10.31577/cai 2022 5 1186.
26. E. Cayirci1, A. Garaga, A. Santana de Oliveira, Y. Roudier. A risk assessment model for selecting cloud service providers. Journal of Cloud Computing: Advances, Systems and Applications (2016), DOI 10.1186/s13677-016-0064-x
27. A Risk Assessment Framework for Cloud Computing. URL: http://eprints.whiterose.ac.uk/95981/
28. ISO/IEC 27001:2022. Information security, cybersecurity and privacy protection. Information security management systems: https://www.iso.org/standard/27001
29. ISO 31000:2018. Risk management: https://www.iso.org/iso-31000-risk-management.html

Received (Надійшла) 23.12.2023
Accepted for publication (Прийнята до друку) 31.01.2024

An expert system for cloud service risk assessment

L. Nikitina, N. Dzheniuk, L. Borysova

Abstract. Modern realities require society to implement digital technologies that are becoming increasingly complex and intelligent. Digitalization is becoming an integral component of all spheres of human activity. The trend of development and economic growth of companies and organizations is cloud technologies, which allow organizations to have flexible costs in the IT sector and regulate them by purchasing access to resources and services from providers instead of purchasing the resources and services themselves. For organizations that make decisions related to the use of cloud services, there are difficulties in evaluating and optimally choosing services and providers, because there are no generally accepted guidelines or procedures for this yet. On the other hand, providers face the problem of ensuring the proper quality of cloud services provided to users. Both providers and users need to have tools that allow them to identify and assess the possible risks of cloud services. One of such tools can be an expert system for evaluating cloud services, the concept of which is considered in this article.

Keywords: Cloud Computing; Cloud Services; Expert System; Fuzzy Inference System; Knowledge Base; Risk Assessment; Vulnerabilities.

УДК 621.395 **doi: 10.26906/SUNZ.2024.1.152**

Д. С. Положий, О. О. Орєхов

Національний аерокосмічний університет імені М. Є. Жуковського «ХАІ», Харків, Україна

МОДЕЛЮВАННЯ ЯКОСТІ ОБСЛУГОВУВАННЯ В АВТОМОБІЛЬНІЙ МЕРЕЖІ ITS

Анотація. Розглянуто якість передачі повідомлень у бездротовій автомобільної мережі, як ключового компонента надійності інтелектуальних транспортних систем (ITS) з акцентом на технології зв'язку між об'єктами автомобільної мережі VANET. Аналіз результатів попередніх досліджень якості зв'язку за стандартом IEEE 802.11p виявили його нездатність підтримувати обслуговування сучасних автомобільних програм. Досліджено особливості передачі даних в мережі за стандартом IEEE 802.11bd, який рекомендовано у останньому звіті європейської CCAM платформи [1] для обміну повідомленнями «транспортний засіб - транспортний засіб» і «транспортний засіб - інфраструктура». Проведено аналітичне моделювання якості обслуговування мережі за стандартом IEEE 802.11bd з використанням марківських ланцюгів. Результати моделювання підтвердили гіпотезу, що протокол передачі повідомлень за стандартом IEEE 802.11bd забезпечує високу якість зв'язку в мережі VANET.

Ключові слова: інтелектуальна транспортна система, автомобільна мережа, аналітична модель, IEEE 802.11 bd, якість, виділений зв'язок, обслуговування мережі.

Вступ та опис проблеми

Посилення інтенсивність транспортного руху на сучасних трасах збільшує небезпеку дорожнього руху і кількість ДТП, ускладнює пересування по великих містах, створюючи «пробки» та підвищує небезпеку для пішоходів. Інтелектуальні транспортні системи (ITS) покликані вирішити проблеми дорожнього руху, і, крім того, сформувати інфраструктуру для транспорту з автоматичним управлінням.

ITS складаються з підсистем мобільного, стаціонарного і верхнього рівня (рис. 1), які обмінюються інформаційними повідомленнями різних стандартів, що пов'язано з різними вимогами до швидкості та дальності передачі сигналу зв'язку.

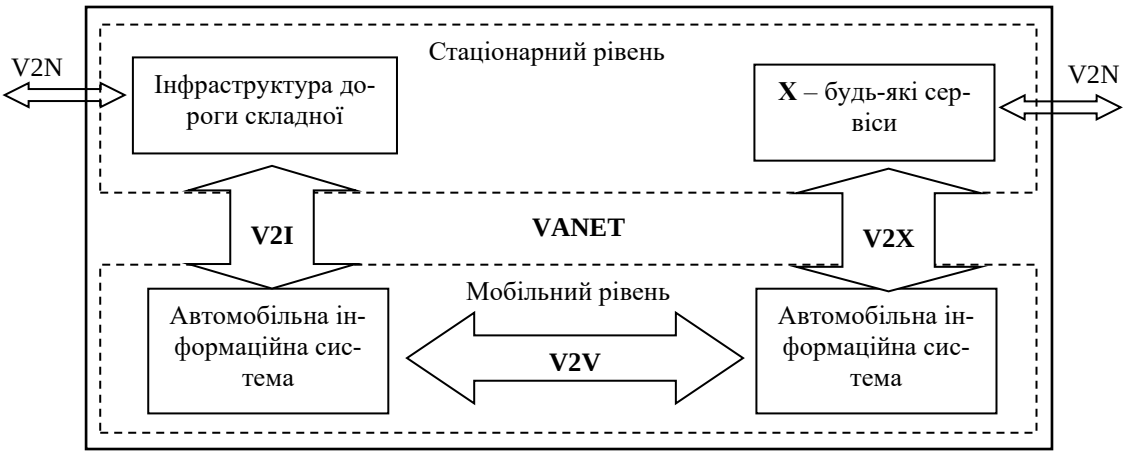


Рис. 1. Структура мобільної транспортної мережі VANET

Мобільний рівень ITS складають розумні транспортні засоби (OBU), які надсилають одне одному повідомлення типу V2V (транспортний засіб – транспортний засіб), створюючи таким чином самоорганізовану мережу для запобігання небезпечним ситуаціям. Крім того, за допомогою зв'язку V2I (транспортний засіб – інфраструктура) здійснюється обмін інформацією з пристроями дорожної інфраструктури (RSU), яка отримує повідомлення з мобільних об'єктів дорожнього руху та оповіщає водіїв (автопілотів) про події на ділянці дороги на інші повідомлення, як то правила дорожнього руху, обмеження швидкості тощо. Всі інфраструктурні і транспортні об'єкти мережі забезпечені крім того зв'язком типу V2N із сервером та стільниковими системами, що складають верхній рівень ITS, який здійснює загальне управляння дорожнім рухом, координацію різних його сегментів.

VANET (Vehicular Ad-Hoc Network) — це технологія, яка забезпечує інтелектуальний зв'язок між мобільними транспортними засобами, заснована на використання виділеного зв'язку малого радіусу дії. Інтеграція VANET із універсальними сенсорними мережами (USN) має великий потенціал для підвищення безпеки та ефективності дорожнього руху. Більшість додатків VANET застосовуються в режимі реального часу і чутливі до затримок, особливо ті, що стосуються безпеки та здоров'я людини. Тому моделювання різних режимів процесу зв'язку для нового стандарту є необхідним і дуже важливим.

Якість зв'язку у мережі VANET має найбільш критичне значення для забезпечення надійності

роботи ITS, оскілки саме від цієї дільниці системи залежатимуть вчасні дії водія по запобіганню ДТП на дорозі. Цей зв'язок повинен відповідати вимогам швидкості та достовірності передачі повідомлень, особливо таких як «екстрене гальмування», «небезпечна ситуація» або «пробка». З іншого боку, у разі повідомлення колони автомобілів засобами VANET, виникають затримки внаслідок малого радіусу дії сигналу, а передача його по ланцюгу V2V викликає додаткові затримки. Отже , ключовою проблемою, яку слід вирішити в бездротових автомобільних мережах є забезпечення якості обслуговування (QoS).

Аналіз останніх досліджень і публікацій

Перспективність і масштабність впровадження проектів інтелектуальних транспортних систем викликає широкий інтерес дослідників та проектувальників – від загальних принципів формування основних функціональних складових системи, структури багаторівневої логіки надання послуг і обслуговування потоків даних [2] до дослідження ймовірності справного стану елементарного фрагмента системи [3]. З'являється також не мало сервісів і додатків для автомобільних інтелектуальних мереж, і вимоги до якості обслуговування їх інформаційних обмінів збільшуються.

У 2010 році для підтримки зв'язку у мережі VANET було запропоновано спеціальний стандарт виділеного зв'язку IEEE 802.11p. Для оцінки та аналізу характеристик зв'язку по стандарту IEEE 802.11p фахівцями було проведено низку досліджень продуктивності каналу [4, 5, 6].

В [4] Yuan Yao із співавторами провели оцінювання продуктивності протоколу IEEE 802.11p MAC (medium access control), який використовується в забезпеченні безпеки транспортних засобів V2V з двома пріоритетними повідомленнями в звичайному середовищі шосе. По результатам проведеного аналізу, автори запропонували покращити продуктивність пріоритетного трафіку за допомогою IEEE 802.11p EDCA для підтримки якості передачі. Однак передача сигналів з нижчим пріоритетом, за спостереженнями авторів, не відповідає вимогам якості обслуговування при передачі у реальному часі.

У [5] Song і співавтори проаналізували продуктивність багатоканального IEEE 802.11p у мережах VANET. Дослідники розглянули декілька особливостей стандарту та багатоканальності за допомогою перемикань у своїй моделі. Були розроблені одновимірні та двовимірні моделі ланцюга Маркова, що створює більшу обґрунтованість оцінки аналізу. Але не враховувався вплив помилок передачі на продуктивність мережі.

В [6] автори виконали моделювання продуктивності сценарію V2I (транспортний засіб – інфраструктура) у загальному та пакетному масштабі системи в каналі передачі даних стандарту IEEE 802.11pWAVE за допомогою марківських ланцюгів.

По отриманим результатам набули висновку, що ймовірність помилки передачі є параметром, який має значний вплив на точність роботи мережі. Відмова від цього параметра в моделі IEEE 802.11p EDCA обмежує її можливості, тому що модель не зможе правильно передбачити продуктивність мережі в схильному до помилок каналі. Збільшення інтенсивності транспортного руху і бітові помилки призводять до зменшення якості обслуговування мережі, що веде до збільшення кількості зіткнень на дорозі.

Отже, дослідження мереж IEEE 802.11p показало, що вони можуть задовольнити вимоги більшості програм безпеки та ефективності трафіку, якщо мережевий трафік помірний.

Але мережі, побудовані на протоколі стандарту IEEE 802.11p навряд чи відповідатимуть більш жорстким вимогам нових програм для автомобільної мережі, які накладатимуть неоднорідні і більш жорсткі вимоги - найсуворіші встановлюють вимогу для максимальної затримки 3 мс, наприклад такі, як додатки автономного керування транспортним засобом та керування колоною.

Вимоги до деяких сучасних транспортних додатків, які використовуються у VANET, були вивчені у [7]. Програми для автомобільної мережі поділяються на чотири категорії: керування транспортним засобом, автоматизоване водіння, автоматизовані датчики та дистанційне водіння. Вимоги до якості обслуговування (QoS) мережею передачі даних для наведених сегментів додатків узагальнено в табл. 1.

Для усунення недоліків стандарту IEEE 802.11p, в 2022 році IEEE запустив новий проект IEEE 802.11bd [8], який є розвитком попереднього. Він визначає модифікації як фізичного рівня IEEE 802.11 (PHY), так і підрівня керування доступом до середовища (MAC) для зв'язку наступного покоління транспортного засобу з усім (V2X) що працює на максимальній обов'язковій швидкості передачі даних в діапазоні 5,9 ГГц (12 Мбіт/с у каналі 10 МГц) і 60 ГГц [8].

Таблиця 1 – **Вимоги до QoS у різних сегментах додатків**

Сегмент використання	Макс. затримка (мс)	Розмір пакета передачі (байт)	Надійність, %	Швидкість передачі даних, (Мбіт/с)	Мін. діапазон сигналу (м)
Керування авто	10-500	50-6000	90-99,99	50-65	80-350
Автоматизоване водіння	3-1, 3-100	300-12000	90-99,99	10-50	360-500
Автоматизовані датчики	3-1, 3-100	1600	90-99,99	10-1000	50-1000
Дистанційне водіння	5	-	99,99	UL:25 DL:1	-

Джерело: складено автором на основі [7].

Особливості використання нового стандарту IEEE 802.11bd досліджено в багатьох роботах [9, 10]. У [10] представлено порівняння стандартів IEEE 802.11bd, IEEE 802.11p і Cellular V2X, показано, що мережі на основі C-V2X і IEEE 802.11bd сприяють зменшенню кількість ДТП у порівнянні з мережами на основі IEEE 802.11p. Важливою особливістю нового стандарту IEEE 802.11bd є техніка зв'язування каналів, яка дозволяє передавати дані в двох суміжних каналах одночасно, що збільшує швидкість передачі даних і, відповідно, зменшує затримки та коефіцієнт втрат пакетів. Але практичне використання систем передач за протоколом IEEE 802.11bd потребує попередніх досліджень.

Постановка завдання

Метою статті є дослідження якості обслуговування каналу зв'язку V2X у мережі VANET, побудованому на стандарті IEEE 802.11bd , із швидкістю передачі даних в діапазоні 5,9 ГГц (12 Мбіт/с у каналі 10 МГц). Для цього проводимо математичне моделювання якості обслуговування по сценарію V2I, яке враховує рівень MAC і обробку сигналів вищого рівня, враховуючи характеристики трафіку трьох основних ITS-сервісів, наведені у табл. 2.

Таблиця 2 – **Параметри сигналів для послуг**

Параметри	Значення для послуг		
	1	2	3
Розмір сповіщення, байт	171	50	200 - 1200
Час між пакетами, с	0,1	0,1	1
Розподіл	Пуассона		
Тривалість періоду увімкнення, с	60		
Тривалість періоду вимкнення, с	20		

Джерело: складено автором на основі [11].

Виклад основного матеріалу

Програми безпеки привернули значну увагу, оскільки вони безпосередньо пов'язані з мінімізацією кількості аварій на дорозі. Категорія безпеки співставна з класом Active Road Safety, який спрямовано на надання послуг поінформованості водія та попередження через програми «Кооперативна обізнаність» (CA), «Кооперативна допомога водіям» (CDA) і «Попередження про небезпеку на дорозі та зіткнення» (RHCW). Активну безпеку на дорозі забезпечують насамперед функції інформування, які передають інформацію водієві під час звичайного водіння, попереджають водія про небезпечні умови на дорозі та ймовірні аварії та активно допомагають водієві уникати загрозливих аварій. Іншими словами, програми, пов'язані з безпекою, відповідають за: інформування, попередження та допомогу. Тому для моделювання у обираємо інфраструктурний прилад, який надає послуги наведених трьох видів.

Послуга 1: Попередження про туманну зону. Покликана попередити водія, що знаходиться поблизу туманної зони, про неминучу небезпеку. Це тип послуг сповіщення про небезпеку на дорозі (RHCN). Ця послуга належить до програми попередження про небезпеку на дорозі та зіткнення (RHCW).

Послуга 2: Міждистанційне вимірювання. Транспортний засіб (OBU) використовує обмін даними датчиків і надсилає вимірювання, пов'язані з відстанню між транспортними засобами. Це допомагає водіям підтримувати безпечну дистанцію між транспортними засобами, щоб гарантувати, що екстрене гальмування не призведе до зіткнень між автомобілями ззаду. Це тип послуги кооперативної системи водіння.

Послуга 3: Попереджувальна подія на дорозі. OBU оснащений фронтальною камерою, яка знімає та надсилає фото при виявленні події попередження про дорогу (поворот дороги, пагорб, швидкість дорожнього знака тощо) тим самим здійснює попередження та фотоповідомлення. Це дозволяє інфраструктурній службі приймати рішення про актуальність події попередження. Це тип служби сповіщень про особливості дороги. Послуга належить до програми попередження про небезпеку на дорозі та зіткнення (RHCW).

Оскільки на якість обслуговування мережі значно впливають варіації та статистичний розподіл трафіку, в моделі потрібно продемонструвати характеристики кожної із служб та виконати моделювання трафіку.

Опис сервісів. Сервіс 1: Попередження для служби зони туману (DEN) Передача повідомлень триває протягом часу перебування станції в зоні дії RSU. Оскільки повідомлення періодично генеруються у разі виявлення події, передача повідомлень відбувається протягом періоду активності (період ON); час, що залишився, є інтервалом часу мовчання (період ВИМК.). Період увімкнення триває до тих пір, поки станція OBU не покине зону RSU або не припинить реєструвати небезпечну подію.

Сервіс 2: Служба міждистанційного вимірювання (CAM). Повідомлення спільної інформації (CAM) періодично передаються кожним транспортним засобом і містять інформацію, зібрану з бортових датчиків. У моделі розглядаємо CAM-повідомлення, що несуть метрику міжстанційного транспортного засобу. Кожну секунду OBU надсилає на RSU повідомлення розміром 50 байт із частотою 10 Гц.

Сервіс 3: Подія попередження про дорогу (CoDM). Коли OBU автомобіля знаходиться в зоні покриття RSU, він надсилає зображення після виявлення події попередження про дорожній об'єкт. Залежно від програми зображення може бути надіслано як одним блоком, так і декількома фрагментами. Попередження про дорожній рух приймає модель ON/OFF; період увімкнення означає час, протягом якого автомобіль надсилає фотопакети, а період вимкнення – це час, протягом якого автомобіль знаходиться в зоні покриття RSU, але не надсилає будь-який пакет. Під час періоду увімкнення OBU щосекунди надсилає повідомлення CoDM розміром від 200 до 1200 байт до RSU протягом часу перебування OBU.

Моделювання пристроїв та мережевого рівня зв'язку. Рівень об'єктів. Приймаємо наступне припущення щодо трьох класів обслуговування:

r – це клас обслуговування (у нас 1, 2, 3), мережевий/транспортний рівень;

λr – інтенсивність для обслуговування r потоку надходження трафіку (незалежні процеси Пуассона);

μr – параметр для послуги r швидкості обслуговування, що відповідає експоненціальному розподілу;

M – кількість транспортних засобів OBU,

ni (i=1,…,M) – загальна кількість пакетів в OBUi в зоні покриття RSU, $n=(n1,n2,\ldots,nM)$,

ρi – навантаження OBUi,

ρir, μir та λir – відповідно навантаження, швидкість обслуговування та середня інтенсивність надходження послуги r що працює на OBUi.

eir – середня кількість відвідувань OBUi за класом обслуговування r.

Процес надходження Пуассона та експоненціальний розподіл швидкості обслуговування дозволяють моделювати систему з трьома чергами M/M/1 на кожному рівні. Кожна черга надає послуги одного класу обслуговування. В результаті моделюємо трафік, який передається верхніми рівнями архітектури за допомогою мереж черги повідомлень, або системи масового обслуговування в реальному часі з пріоритетами (СМО). Функція розподілу ймовірностей стаціонарного стану для однокласової (r) незалежної від навантаження відкритої СМО визначається добутком розподілу ймовірностей стаціонарного стану ізольованих черг визначається за формулою:

$$F(n)=\prod_{i=1}^{M} pi(ni),$$

$$pi(ni)=(1-\rho i)ni!\prod_{i=1}^{M}\frac{1}{nir!}\left(\frac{\lambda r\cdot eir}{\mu r}\right)^{nir}.$$

Для кожного класу обслуговування, що виконується на пристрої i, отримуємо параметри якісті обслуговування на рівні засобів/мережі за формулами:

– середня швидкість обслуговування: $Dir=\lambda ir$

– середня кількість пакетів:

$$Lir=\frac{\rho ir}{1-\rho i};$$

– середній час перебування:

$$Tir=\frac{Lir}{Dir};$$

– середній час очікування:

$$Wir=Tir-\frac{1}{\mu ir}.$$

Такий аналіз продуктивності дозволяє оцінити потік інформаційних пакетів на верхніх рівнях і визначає вхідні дані нижнього рівня.

Моделювання радіорівня MAC. Після обробки засобами та мережевим рівнем пакети послуг передаються на рівень MAC радіозв'язку. Щоб реалізувати диференціацію послуг, встановлюються пріоритети пакетам послуг. Зіставимо три служби з категоріями доступу EDCA наступним чином:

- служба попереджень містить критичну інформацію, тому віднесена до категорії доступу з найвищим пріоритетом AC_VO;
- служба вимірювання генерує важливі дані та відображається на AC_BE; часта передача повідомлень CAM долає ненадійність AC_BE;
- служба попередження про дорогу зіставлена з AC_BK.

Параметри EDCA, використані в математичному дослідженні та моделюванні, наведено в табл. 3.

Таблиця 3 – **Параметри категорій доступу EDCA**

ITS Сервіси	EDCA категорії доступу	Cwmin		Cwmax		AIFSN	
		CCH	SCH	CCH	SCH	CCH	SCH
Дорожня подія	AC_BK	15	15	511	511	9	7
Вимірювання міжвідстаней	AC_BE	7	15	15	511	6	3
Попередження про туманну зону	AC_VO	3	3	7	7	2	2

Джерело: складено автором на основі [11]

Моделювання радіорівня MAC виконуємо за допомогою мультікласу M/GI/1 (з трьома класами обслуговування), який приймає політику пріоритету без виключення для кожного класу обслуговування. Пакети, що належать до класу обслуговування, відповідають розподілу Пуассона. Служба дотримується загального закону незалежного прибуття.

Для кожної послуги r позначимо:

E[Nr] – кількість пакетів, що очікують в черзі,

E[Xr] – середній час обслуговування,

E[*Wr*] – середній час очікування;

E[*Rr*] – незавершені процеси в черзі очікування.

За формулою Поллачека-Хінчіна [12] отримаємо основні параметри якості обслуговування для оцінки системи невипереджувального пріоритету M/GI/1 для кожного класу обслуговування *r*:

– середня кількість пакетів *E*(*Nr*):

$$\rho r + \frac{\rho^2 r\left(1+C^2\right)}{2\left(1-\rho r\right)};$$

– середній час перебування *E*(*Sr*):

$$E(Sr) = E(Xr) + E(Wr);$$

- середній час очікування *E*(*Wr*) :

$$\frac{E\left(W0\right)}{\left(1-\rho r\right)\left(1-\sum_{r=1}^{3}\rho r\right)}; \qquad (1)$$

де *C* – коефіцієнт варіації; *E* [*W*0] – середній час очікування в черзі.

Середній час очікування не враховує час відстрочки, тому модифікували вираз (1), щоб охопити ефект відставання:

$$\frac{E\left(W0\right)}{\left(1-\rho r\right)\left(1-\sum_{r=1}^{3}\rho r\right)} + \sum_{i=0}^{cw\min} p(backoffr = i)\ .$$

Оскільки вікно відстрочки є цілим випадковим значенням і дотримуючись [1, CW +1] в рівномірному розподілі розмірів вікна, з Cw Є [Cwmin,Cwmax], ймовірність *P* відобразити вікно відстрочки визначається як:

$$P\left(backoff\ = i\right) = \frac{1}{Cwmin\ +1}\ .$$

Щоб оцінити якість обслуговування мережі V2I, що містить OBU, які запускають три служби, порівнюємо аналітичні результати, отримані з попереднього математичного моделювання, з моделюванням за сценарієм.

Сценарій моделювання. Моделювання проведено за допомогою сімулятора OpenModelica. Змодельована топологія – це односмугове шосе протяжністю 5 км. Досліджуємо один RSU із зоною покриття (радіусом) 200 м. Транспортні засоби рухаються за випадковою моделлю в одному напрямку із середньою швидкістю 20 км/год., що відповідає ситуації затору. За оцінками, транспортний засіб залишатиметься в зоні покриття пристрою RSU 2,5 хвилини.

Кожен об'єкт OBU запускає три змодельовані служби:

- служба попередження (повідомлення DEN із середньою швидкістю надходження Пуассона λ1= 10 повідомлень/с);

- служба вимірювання (повідомлення CAM із середньою швидкістю надходження Пуассона λ2 =10 повідомлень/с;

- подія попередження про дорогу (повідомлення CoDM) із середньою швидкістю надходження Пуассона λ3= 1 повідомлень/с).

Послуги зіставляються з категоріями доступу EDCA згідно таблиці 5 і представляють моделі трафіку, визначені вище. Передбачені послуги генерують пакети, які надсилаються по каналу обслуговування SCH N° 176, з постійною моделлю розповсюдження. Швидкість передачі 12 Мбіт/с контролюється алгоритмом менеджера швидкості, і враховується посилення -10 дБ. Кількість транспортних засобів змінюється в діапазоні [5…100] з кроком 5.

Досліджуємо максимальну кількість транспортних засобів у зоні покриття RSU та здатності RSU для передачі всіх повідомлень. Вимірюмо середню швидкість втрат пакетів і затримки. На рис. 2 показано порівняння середніх показників втрати пакетів у залежності від кількості транспортних засобів.

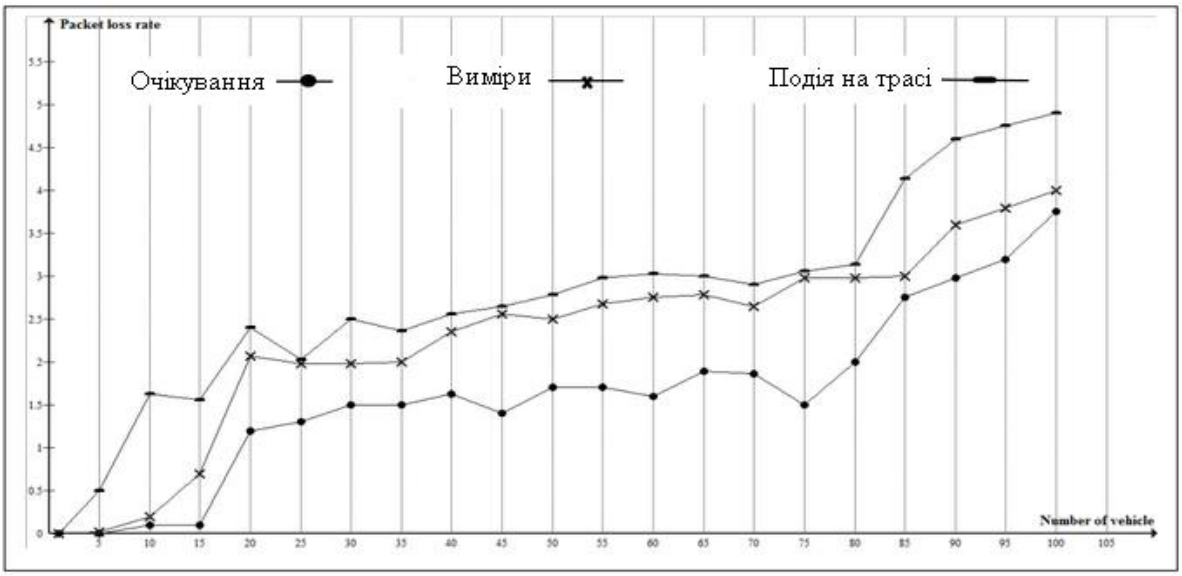


Рис. 2. Залежність середніх показників втрати пакетів від кількості транспортних засобів

Кожна крива пов'язана з однією з послуг: попередження, вимірювання та попереджувальна подія на дорозі. Швидкість втрати пакетів помірно зростає, коли щільність трафіку коливається від 20

до 80 транспортних засобів, але крім цього швидкість значно зростає для всіх служб додатків. Чим більше транспортних засобів пов'язано з одним RSU, тим частіше транспортні засоби відчувають процес відставання, та тим вищий рівень зіткнень. Рівень втрати пакетів у службі попередження про дорогу є вищим, ніж у вимірюванні, яке є вищим, ніж у попереджувальному повідомленні. Головним чином це пов'язано з різними рівнями пріоритетності послуг. Виявлено, що коли кількість транспортних засобів змінюється в діапазоні [5...30], затримка майже однакова для обох послуг.

Однак воно значно збільшується для служби вимірювання, коли номер транспортного засобу перевищує 45. Затримка попередження більш регулярна в діапазоні [5...80]. Цей результат підтверджує продуктивність EDCA, яка успішно визначає пріоритет попередження над послугою вимірювання. Фактично, зміна відставання є лінійною функцією навантаження дорожнього руху для різних діапазонів зон (300 м, 500 м і 900 м).

Висновки

У роботі проаналізовано структуру бездротової автомобільної мережі інтелектуальної транспортної системи та фактори впливу на якість її обслуговування. Доведено, що найбільше на якість обслуговування в ITS впливають технології зв'язку. Стандарт IEEE 802.11bd спрямований на підвищення надійності передачі, збільшення пропускної здатності та дальності передачі в порівнянні зі стандартом IEEE 802.11p. Проведене моделювання обслуговування інфраструктурним приладом транспортного потоку за допомогою мобільної транспортній мережі підтвердило диференціацію якості обслуговування послуг за різними пріоритетами і дотримання критичного характеру служби попередження для інтенсивного дорожнього руху.

Перспективами подальших досліджень є розвиток моделі обслуговування в ITS в напрямку збільшення кількості послуг, визначення оптимальної кількості інфраструктурних приладів для якості обслуговування, яка забезпечує безпечний дорожній рух.

СПИСОК ЛІТЕРАТУРИ

1. Final report of the Single Platform for Open Road Testing and Pre-deployment of Cooperative, Connected and Automated and Autonomous Mobility Platform (CCAM platform). URL: https://transport.ec.europa.eu/document/download/15d0f5a7-73cd-48f0-83d3-1c7a160d5854_en?filename=Final%20Report-CCAM%20Platform.pdf
2. Положий Д. С., Орєхов О. О. Інтелектуальні системи автомобільної безпеки на основі хмарних архітектур // Системи управління, навігації та зв'язку. 2023. № 4. С. 91-95
3. Борисова Л. В., Закора О. В., Фещенко А. Б. Розробка імовірнісної моделі елементарного фрагмента відомчої інформаційно-телекомунікаційної мережі // Проблеми надзвичайних ситуацій, 2020. № 1. С. 34-43
4. Yuan Yao, Yujlao Hu, Gang Yang, Xingshe Xhou. On MAC Access Delay Distribution for IEEE 802.11p Broadcast in Vehicular Networks [2019]. DOI: 10.1109/ACCESS.2019.2946989
5. Song C., Performance analysis of the ieee 802.11p multichannel mac protocol in vehicular ad hoc networks, Sensors 17 (12) [2017] 2890.
6. Jose R. Gallardo, DimitriosMakrakis, Hussein T.Mouftah. Mathematical Analysis of EDCA's Performance on the Control Channel of an IEEE 802.11pWAVE Vehicular Network [2017]
7. Study on Enhancement of 3GPP Support for 5G V2X Services (v16.2.0 Release 16) [2018].
8. 802.11bd-2022 - IEEE Standard for Information Technology--Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks--Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Next Generation V2X
9. Yacheur B. Y., Ahmed T., Mosbah M., "Implementation and assessment of IEEE 802.11 BD for improved road safety", Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf. (CCNC), pp. 1-6, Jan. 2021.
10. Gaurang Naik; Biplav Choudhury; Jung-Min Park. IEEE 802.11bd & 5G NR V2X: Evolution of Radio Access Technologies for V2X Communications [2019] DOI: 10.1109/ACCESS.2019.2919489
11. IEEE Trial-Use Standard for Wireless Access in Vehicular Environments (WAVE)—Multi-channel Operation.
12. Литвинов А.Л. Розробка та дослідження ймовірнісних моделей оцінювання якості інформаційно-управляючих систем // Комунальне господарство міст, т, 2016, вип.. 126. С. 22-27

Received (Надійшла) 15.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Simulation of communication quality in the vehicle its network

Denys Polozhyi, Oleksandr Oriekhov

Abstract. The quality of message transmission in a wireless automotive network is considered as a key component of the reliability of intelligent transport systems (ITS) with an emphasis on communication technology between VANET automotive network objects. Analysis of the results of previous studies of the quality of communication according to the IEEE 802.11p standard revealed its inability to support the maintenance of modern automotive applications. The peculiarities of data transmission in the network according to the IEEE 802.11bd standard, which is recommended in the latest report of the European CCAM platform [1] for the exchange of "Vehicle-Vehicle" and "Vehicle-Infrastructure" messages, have been studied. Analytical modeling of network service quality according to the IEEE 802.11bd standard using Markov chains was carried out. The simulation results confirmed the hypothesis that the IEEE 802.11bd message transmission protocol provides high quality communication in the VANET network. Keywords: intelligent transport system, car network, analytical model, IEEE 802.11 bd, quality, dedicated communication, network maintenance.

Keywords: intelligent transport system, automobile network, analytical model, IEEE 802.11 bd, quality, dedicated communication, network maintenance.

УДК 004.032.26 **doi: 10.26906/SUNZ.2024.1.158**

С. І. Шаповалова, А. Ю. Софієнко

Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ, Україна

ЦИФРОВІ ПРЕДСТАВЛЕННЯ TELEGRAM-КАНАЛІВ

Анотація. Предметом дослідження цієї статті є цифрові представлення текстових інформаційних ресурсів на прикладі Telegram-каналів. Мета роботи – визначити оптимальний для подальшої тематичної класифікації метод формування цифрових представлень Telegram-каналів. У статті вирішуються наступні завдання: означення підходів до формування вхідного вектору; визначення етапів обробки текстових даних для цифрового представлення Telegram-каналу; створення датасету цифрових представлення Telegram-каналів; розмітка датасету для розв'язання задачі класифікації; визначення гіперпараметрів оптимальних моделей класифікації. Отримано такі результати: створений датасет цифрових представлень Telegram-каналів, сформованих на основі мережі SBERT, за трьома підходами: агрегація векторів публікацій, конкатенація ключових слів за методом TF-IDF та поєднання перших двох підходів; визначено, що підхід конкатенації ключових слів за методом TF-IDF та поєднання перших двох підходів до формування цифрових представлень Telegram-каналів на основі текстових публікацій є найбільш ефективним для подальшої класифікації за тематикою; визначено оптимальні за точністю гіперпараметри моделей тематичної класифікації: Logistic Regressio та нейромережі глибокого навчання. Перспективним напрямком подальших досліджень є оцінювання застосування запропонованих цифрових представлень до задач кластеризації та пошуку.

Ключові слова: обробка текстів природною мовою, BERT, тематична класифікація повідомлень, representation learning.

Вступ

Онлайн-соціальні мережі (OSN) становлять життєво важливий аспект сучасної комунікації, який є актуальним у повсякденному житті. Онлайн-платформи та соціальні мережі стали ключовим джерелом інформації для більшої частини населення світу. Невпинне зростання кількості джерел та обсягів інформації призводить до нових проблем. Перенасиченість великою кількістю інформації зробила пошук, фільтрацію, пошук релевантних джерел інформації, їх структуризацію та аналіз складним та ресурсовитратним завданням. Оскільки публікації в соціальних мережах – це сильно розріджені текстові дані з великою кількістю шумів, для подальшої роботи з ними необхідний інструмент якісного вилучення "дистильованих" фактів.

Сучасним підходом до аналізу даних є застосування алгоритмів машинного навчання, які приймають на вхід цифрові величини – вектори ознак, якість та чистота яких напряму впливає на результат роботи алгоритмів. Згідно з дослідженням компанії Gradus Research Company [1], яке проводилось на замовлення Національної суспільної телерадіокомпанії України, понад 55% українців отримують інформацію з месенджерів та соціальних мереж. Telegram – лідер серед месенджерів – 89% співвітчизників користуються ним для перегляду новин.

Аналіз останніх досліджень і публікацій

Соціальні мережі стали не тільки потужними інструментами для спілкування та обміну інформацією, а й об'єктом інтенсивних досліджень в галузі обробки природної мови на основі машинного навчання. Лише за останні декілька років було проведено велику кількість досліджень, пов'язаних з використанням інформації з онлайн соціальних мереж.

Довгий час для створення цифрових представлень інформації природною мовою використовувалися класичні алгоритми Bag of Words та TF-IDF, що вираховували лише частоту використання слів чи словосполучень (англ. N-gram) в документах, при цьому втрачаючи семантичну інформацію. Проте в останнє десятиліття спостерігається стрімкий розвиток більш складних та контекстно-орієнтованих методів обробки природної мови. На відміну від традиційних статистичних методів, до розуміння семантики та контексту використовуються нейронні мережі, що створюють представлення (англ. embedding) слів та текстів з врахуванням контекстної інформації.

Починаючи з 2013 року, для створення векторів слів було представлено такі підходи:

- Word2Vec (Google, 2013) [2],
- GloVe (Stanford University, 2014) [3],
- FastText (Facebook AI, 2016) [4].

Ці алгоритми використовують неглибокі мережі для побудови представлень для слів. З розвитком глибоких нейронних мереж в 2018 році дослідниками з Google AI було представлено архітектуру BERT [5], що здатна оперувати реченнями та фрагментами тексту, а не окремими словами. Через рік розроблено SBERT (Sentence-BERT) [6] – модифікацію BERT, що використовує для навчання сіамську та триплет мережу, представлення якої можна порівнювати за косинусної відстанню, що є важливим в задачах пошуку та кластеризації.

Дослідники використовують різні архітектури та моделі, такі як глибокі нейронні мережі та методи графового аналізу, для обробки даних з таких платформ як Twitter, Facebook, Instagram та інші. В роботі [7] було застосовано алгоритм Spherical k-means для кластеризації distilBERT-представлень з метою визначення основних тем дезінформації, виявлення ключових трендів. Було отримано високу точність в оцінках на наявному датасеті. Подібним чином в роботі [8] були використані ембедінги, для класифікації акаунтів ботів та людей в мережі Twitter.

З моменту впровадження архітектури з основою BERT у сфері обробки природної мови, цей підхід

отримав значний успіху завдяки своїй здатності до розуміння контексту та синтаксичних залежностей. Однак, незважаючи на великий потенціал BERT, її вхід базується на токенах та має фіксоване обмеження на розмір, що може суттєво впливати на використання мережі в конкретних сценаріях. Архітектура BERT обмежена за кількості вхідних токенів, які можуть бути оброблені моделлю. У класичній версії моделі ця обмеженість складає 512 токенів. В розширених версіях, таких як BERT-large, вхід розширено до 1024 токенів, проте цього недостатньо для опрацювання об'ємних текстових даних.

Таким чином, для оптимального використання BERT-моделей та подолання їх обмежень потрібне вдосконалення методів створення цифрових представлень. Постає необхідність у розробці методів агрегації або скорочення тексту, щоб зберегти важливу інформацію та забезпечити ефективне використання BERT на великих корпусах тексту.

Однак, важливо врахувати, що агрегація тексту може призвести до втрати семантичної інформації, тому необхідно вдосконалення методів ембедінгу для збереження значущості текстового контексту під час агрегації.

Отже, для побудови цифрових представлень Telegram-каналів необхідне подолання обмежень розміру входу нейронної мережі SBERT.

Постановка задачі

Метою статті є визначення оптимального для подальшої тематичної класифікації методу формування цифрових представлень Telegram-каналів.

Завдання:

– означити підходи до формування цифрових представлень Telegram-каналів на основі мережі SBERT;

– визначити етапи обробки текстових даних для цифрового представлення Telegram-каналу;

– створити датасет цифрових представлення Telegram-каналів;

– розмітити датасет для розв’язання задачі класифікації;

– експериментально визначити гіперпараметри оптимальних моделей класифікації.

Вхідною інформацією є необроблені текстові дані публікацій в Telegram-каналах українською та російською мовами.

Результатом є цифрове представлення, яке придатне для подальшої обробки та вирішення завдань із застосуванням машинного навчання для класифікації, фільтрації, агрегації, категоризації, рекомендації.

Модель штучного інтелекту - навчена нейронна мережа SBERT, ресурс HuggingFace [9].

Подолання обмеженості розміру входу моделі SBERT для формування цифрових представлень Telegram-каналів здійснюється підбором оптимальних параметрів за критерієм точності тематичної класифікації відповідних методів за такими підходами:

– агрегація векторів публікацій;

– вектор конкатенації TF-IDF ключових слів;

– агрегація векторів публікацій та ключових слів.

Результати досліджень

1. Підходи до формування цифрового представлення Telegram-каналів. Модель SBERT має ліміт входу 512 або 1024 токенів. Для подолання цього обмеження може бути використано декілька підходів.

1. *Агрегація векторів публікацій*: входом моделі є публікація. Замість подання моделі для обробки всього корпусу текстів Telegram-каналу використовуються окремі публікації. Кожна з них окремо подається до SBERT. Таким чином для всіх текстів отримуються вектори прихованих станів. Обчислюється середнє значення цих векторів, що дозволяє узагальнити та зберегти суттєву інформацію з різних повідомлень в одному компактному представленні.

2. *Вектор конкатенації TF-IDF ключових слів*: входом моделі є текст -конкатенація ключових слів. Такий метод передбачає виділення важливих термінів у тексті за допомогою TF-IDF та об'єднання цих термінів у вектор. Цей метод дозволяє враховувати суттєві та репрезентативні елементи тексту, винятково акцентуючи увагу до ключових словосполучень ресурсу.

3. *Агрегація векторів публікацій та ключових слів*: входом моделі є публікації та конкатенація ключових слів. Цей підхід враховує ключові терміни та контекст окремих публікацій, що можуть спільно вносити вагомий внесок у розуміння тексту моделлю SBERT.

2. Конвеєр обробки текстових даних Telegram-каналу. Створення числового представлення Телеграм-каналу здійснюється за декілька етапів і потребує вирішення як загальних так і специфічних задач обробки тексту. На рис. 1 представлено схему конвеєра обробки тексту (англ. data pipeline) для визначення характеристичних ознак тексту.

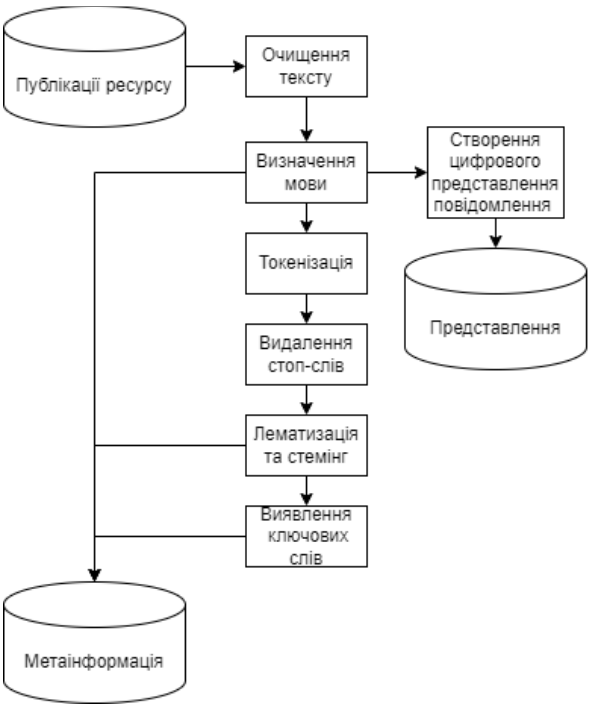


Рис. 1. Конвеєр обробки тексту

Очищення вхідного тексту. Першим етапом обробки публікації з ресурсу є очищення від шуму: тегів, спецсимволів, веб-посилань. Цей етап дозволяє подальшим алгоритмам та моделям краще концентруватися на суттєвому змісті.

Фільтрація на основі визначення мови публікації. Задачею цього етапу є виокремлення повідомлень російською та українською мовами. Кожна мова має специфічні особливості: граматику, лексику та правила, які впливають на обробку тексту. Визначення мови допомагає вибрати відповідні методи та алгоритми для обробки тексту конкретною мовою, що забезпечує більш точні та ефективні результати. На цьому етапі для ідентифікації мови було використано бібліотеку fastText від Facebook's AI Research, що здатна визначити ISO код мови тексту з-поміж 176 доступних.

Очищений від шуму текст та його мова зберігаються як метаінформація публікації. Оскільки основними мовами набору даних дослідження є українська та російська, публікації іншими мовами далі в обробку не йшли.

Токенізація та видалення стоп-слів. Наступним етапом конвеєру обробки є розбиття слів на базові елементи – токени. В роботі для токенізації було використано Python бібліотеку NLTK[10]. З упорядкованої колекції токенів відфільтровуються стоп-слова — токени, що не несуть суттєвої інформації. Список стоп-слів обирався залежно від мови тексту. За основу було взято наявні списки з бібліотеки NLTK. Оскільки на цьому ресурсі не представлено українську та недостатньо представлено російську, колекція стоп-слів була доповнена прикладами з ресурсів [11, 12] . Також додано слова з власних спостережень.

Лематизація, стемінг та морфологічний аналіз. Оскільки може існувати велика кількість форм одного й того ж слова, для полегшення аналізу тексту слова приводяться до базової форми. В результаті декілька слів в різній формі трактуються як єдине представлення. Для отримання початкової форми слів було використано бібліотеку Python pymorphy2 [13]. Однією з функцій цієї бібліотеки є здатність визначати морфологічну характеристику слова, яку було використано для виявлення власних назв з метою доповнення ключовими словами метаінформації про публікацію.

Визначення ключових слів ресурсу. Для швидкої оцінки тематики ресурсу необхідно виокремити ключові слова, які відрізняються від загальних та мають важливу інформаційну цінність. Для визначення ключових термінів в контексті всього ресурсу було використано статистичний метод TF-IDF (Term Frequency-Inverse Document Frequency) з бібліотеки scikit-learn [14]. На цьому етапі метаінформація також доповнюється ключовими словами. Для цього визначається топ-100 слів з найвищим значення TF-IDF метрики, яка відображає важливість слова в поточній публікації.

3. Створення датасету. Наявність якісного датасету є одним з ключових факторів отримання хорошого результату машинного навчання.

На сьогоднішній день існує декілька публічно доступних наборів даних з Telegram:

1) багатомовний Pushshift Telegram [15];

2) перською мовою Dataset-for-teenagers-chat-in-Telegram-groups [16];

3) болгарською мовою TRACES Bulgarian Telegram Dataset [17];

4) на тематику криптовалют Crypto telegram groups [18].

Проте існуючі дані непридатні для поставленої задачі через тематичну спрямованість та мови представлення. Тому було створено датасет з текстових даних та посилань з відкритих Telegram-каналів українською та російською мовами. Для збору датасету було розроблено кравлер (англ. crawler) - спеціальну програму, яка автоматично обходила канали за стратегією в ширину та завантажувала дані. Процес збору включав в себе переходи за посиланнями на канали, вилучення публікації та метаінформації, збереження їх до бази даних. В результаті було створено датасет із 9753 Telegram-каналів.

В подальшому ці дані були оброблені та очищені з метою видалення дублікатів, а також непотрібних даних з набору. Процес відбувався за етапами, наведеними на рис. 1.

4. Розмітка датасету. Зібраний датасет є нерозміченим, тобто він не містить в собі відомостей про приналежність кожного спостереження до конкретного класу або категорії. Для отримання необхідних міток було застосовано класи тематики Telegram-каналів з ресурсу tgstat.com [19].

Використовуючи цей ресурс, було отримано мітки 38 класів для 5691 з 9753 каналів. Класи решти Telegram-каналів в створеному датасеті ресурс tgstat.com не надає. Перелік та розподіл міток зображено на рис. 2.

У процесі обробки та аналізу даних виявлено, що деякі мітки включали одні й ті ж класи даних, що призводило до небажаних розбіжностей. Так, категорії "Новини", "Політика" та "Блоги" частково перетинаються, що негативно впливає на якість роботи алгоритмів.

З метою покращення якості та уніфікації датасету було виключено такі амбівалентні мітки з вхідного набору даних. Розподіл класів та категорій оновленого датасету зображено на рисунку 3.

Таке оновлення спряло зменшенню надмірної складності даних та усуненню можливої багатозначності при прогнозуванні на моделі штучного інтелекту.

Діаграма на рис. 3 демонструє, що розподіл прикладів по класах став більш збалансованим, однак все ще нерівномірний. Це може призвести до проблеми коректності класифікації. Модель може бути схильною до прийняття більшого класу як "переважаючого", і неадекватно розрізняти менший клас.

Для вирішення проблеми незбалансованості на етапі навчання було визначено ваги класів. Вони були представлені масивом коефіцієнтів для обчислення помилки в процесі навчання.

Гіпотетично модель має краще адаптуватись до різниці у кількості прикладів для кожного класу і покращити якість прогнозів. Ваги було встановлено зворотньо-пропорційно до частоти класів у навчальному наборі.

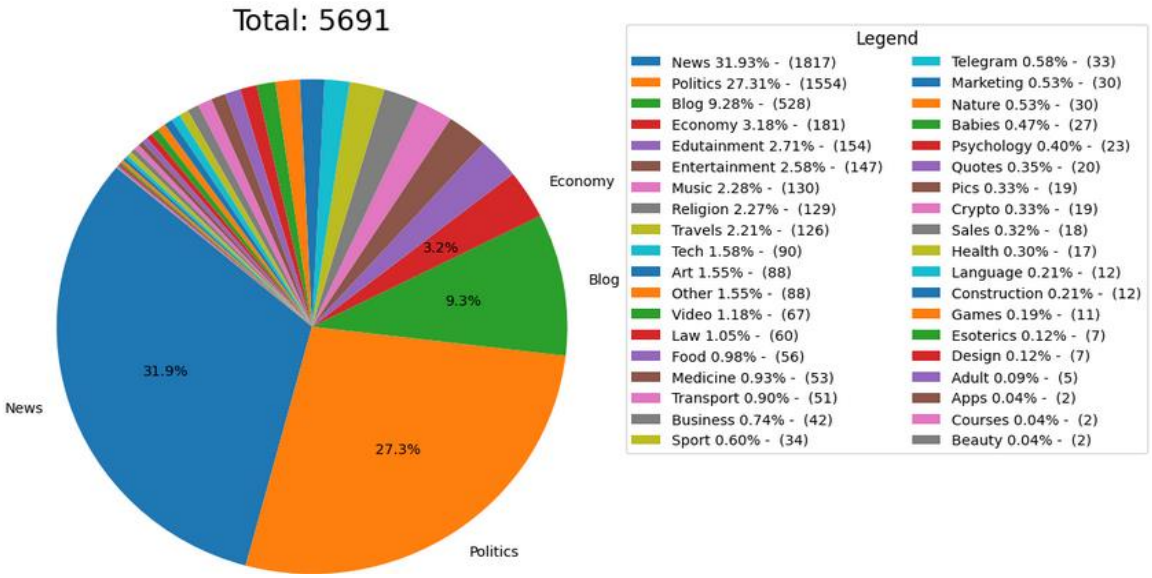


Рис. 2. Кругова діаграма розподілу класів розміченої частини набору даних

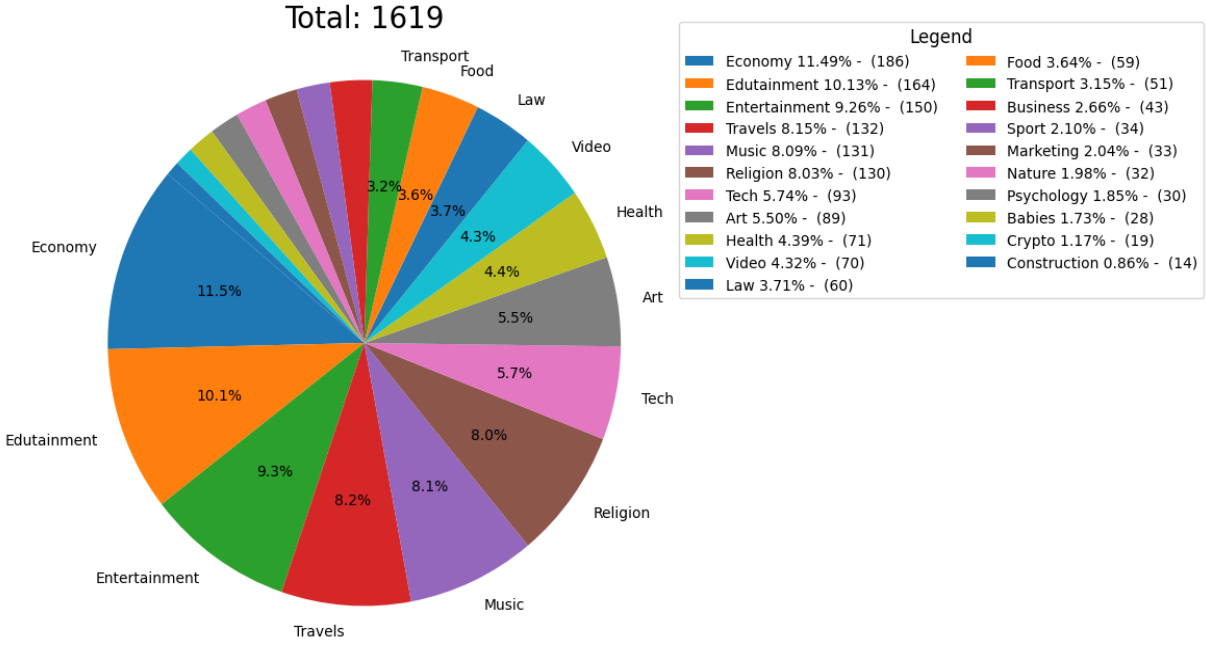


Рис. 3. Кругова діаграма розподілу класів після видалення амбівалентних категорій

Гіпотетично модель має краще адаптуватись до різниці у кількості прикладів для кожного класу і покращити якість прогнозів. Ваги було встановлено зворотньо-пропорційно до частоти класів у навчальному наборі.

5. Постановка задачі обчислювальних експериментів. Задача – мультикласова класифікація Telegram-каналів за тематикою (предметом обговорення) на основі базових підходів штучного інтелекту.

Варіанти цифрових представлень Telegram-каналів:

- агрегація векторів публікацій;
- вектор конкатенації TF-IDF ключових слів;
- агрегація векторів публікацій та ключових слів.

Навчальна вибірка Telegram-каналів та їх представлень становила 1295 каналів, тестова - 324 канали.

Для проведення класифікації були використані два основних набори моделей:

- класичні алгоритми машинного навчання: Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors та Support Vector Classifier (SVC), реалізації яких представлено в бібліотеці Python - scikit-learn.

- повнозв'язні нейронні мережі, реалізовані з використанням фреймворку Keras та TensorFlow [20].

Застосування обох підходів дозволило дослідити та порівняти ефективність методів глибокого навчання та традиційних алгоритмів машинного навчання для задачі класифікації тематики Telegram-каналів.

6. Пошук гіперпараметрів моделей класифікації. Пошук оптимальних гіперпараметрів є важливою складовою процесу налаштування моделей машинного навчання, метою якої є досягнення найкращої продуктивності. Гіперпараметри визначають структуру та математичні параметри поточної моделі. Наприклад, для нейромережі – це кількість

шарів, кількість нейронів в кожному шарі, швидкість навчання, функції активації та інші.

В табл. 1 наведено значення гіперпараметрів моделей за класичними алгоритмами машинного навчання, комбінації яких було застосовано для визначення оптимальних значень поточної моделі.

Гіперпараметрами моделювання нейромережі обрано кількість прихованих шарів та функції активації нейронів за шарами.

Іншими параметрами моделювання були:

- оптимізатори процесу навчання;
- розмір батчу;
- наявність вагів класів.

Пошук оптимальних значень гіперпараметрів нейромережі відбувався шляхом автоматизованого перебору. Всього було перевірено 6048 комбінацій гіперпараметрів для 3 цифрових представлень Telegram-каналів.

В табл. 2 наведено всі параметри моделювання нейромережі та їх значення.

Таблиця 1 – **Гіперпараметри моделей машинного навчання**

Модель	Назва параметрів	Значення
Logistic Regression	C	0.001, 0.01, 0.1, 1, 10, 100
Decision Tree	max_depth	None, 10, 20, 30, 40
	min_samples_split	2, 5, 10
Random Forest	n_estimators	50, 100, 200
	max_depth	None, 10, 20, 30
	min_samples_split	2, 5, 10
K-Nearest Neighbors	n_neighbors	3, 5, 7, 9
	weights	uniform, distance
SVC (Support Vector Classifier)	C	0.1, 1, 10
	kernel	linear, rbf

Таблиця 2 – **Гіперпараметри моделей нейронної мережі**

Параметри моделювання	Значення
Ваги класів (class_weights)	False, True
Кількість прихованих шарів (hidden layers)	[], [64], [128], [256], [512] [512,256,128, 64], [256,128, 64], [128, 64]
Функції активації (Hidden layers activations)	relu, tanh, leakyrelu, sigmoid, elu, prelu
Розмір батчу (batch_sizes)	[256, 128, 64]
Оптимізатори процесу навчання (optimizers)	Adam, SGD, RMSprop

7. Результати класифікації моделями з оптимальними гіперпараметрами. На всіх моделях з визначеними оптимальними гіперпараметрами було розв'язано задачу тематичної класифікації.

В табл. 3 наведено результати класифікації за метриками accuracy, precision, recall, які отримано відповідною моделлю (Classifier) з оптимальними значеннями гіперпарамтрів (Best Parameters) та типом цифрового представлення (Embeding Type).

Найкращі результати за класичними алгоритмами машинного навчання було отримано моделлю Support Vector Classifier з параметром регуляризації C = 10 та ядром радіальної базисної функції (RBF) на представленнях, створених агрегацією векторів публікацій.

В табл. 4 представлено найкращі комбінації нейромережевих моделей та цифрових представлень Telegram-каналів (Embeding Type) та використання вагів класів:

TRUE – наявність коефіцієнтів класів при визначенні помилки в процесі навчання,

FALSE - відсутність).

Таблиця 3 – **Результати класифікації телеграм каналів моделями машинного навчання**

Classifier	Best Parameters	Embeding Type	Accuracy	Precision	Recall
Logistic Regression	C: 0.1	Concatenated	0.783	0.789	0.708
Decision Tree	max_depth: 30; min_samples_split: 2	Concatenated	0.561	0.453	0.461
Random Forest	max_depth: None; min_samples_split: 5; n_estimators: 200	Concatenated	0.774	0.737967	0.633
K-Nearest Neighbors	n_neighbors: 9; weights: distance	Concatenated	0.759	0.722	0.669
SVC	C: 10; kernel: 'rbf'	Concatenated	0.777	0.734	0.699
Logistic Regression	C: 0.1	TfIdfVector	0.722	0.724	0.680
Decision Tree	max_depth: 40; min_samples_split: 10	TfIdfVector	0.389	0.354	0.312
Random Forest	max_depth: None; min_samples_split: 2, n_estimators: 200	TfIdfVector	0.688	0.744	0.545
K-Nearest Neighbors	n_neighbors: 9; weights': 'distance	TfIdfVector	0.710	0.684	0.624
SVC	C: 10; kernel: rbf	TfIdfVector	0.728	0.735	0.681
Logistic Regression	C: 0.1	MessagesMean	0.787	0.762	0.683
Decision Tree	max_depth: 30; min_samples_split: 2	MessagesMean	0.531	0.408	0.409
Random Forest	max_depth: 20; min_samples_split: 2, n_estimators: 200	MessagesMean	0.756	0.712	0.611
K-Nearest Neighbors	n_neighbors: 7; weights: distance	MessagesMean	0.744	0.705	0.647
SVC	C: 10; kernel: rbf	MessagesMean	0.762	0.749	0.674

Таблиця 4 **–Результати класифікації телеграм каналів нейронними мережами**

Model	Embeding Type	Class Weights	Accuracy	Precision	Recall	Auc	Prc
Adam-[64]-prelu-64	MessagesMean	FALSE	0.981	0.866	0.701	0.973	0.852
Adam-[128]-tanh-256	TfIdfVector	FALSE	0.976	0.822	0.642	0.954	0.776
Model-Adam-[64]-sigmoid-128	Concatenated	FALSE	0.983	0.865	0.756	0.980	0.8611

На другому наборі моделей найкращі результати по всіх метриках були досягнуті нейронною мережею з одним прихованим шаром на 64 нейрони, які використовували функцію активації Sigmoid та цифрових представленнях створених агрегацією векторів публікацій. Під час навчання цієї моделі використовувався оптимізатор Adam з налаштуваннями за замовчуванням. На вхід моделі цифрові представлення подавались в батчах, де кожен батч мав розмірність 128 елементи. Звідти позначка моделі Adam-[64]-sigmoid-128.

На рис. 4 представлено залежність значень метрик точності від епох навчання моделі Adam-[64]-sigmoid-128 з найкращим кінцевим результатом.

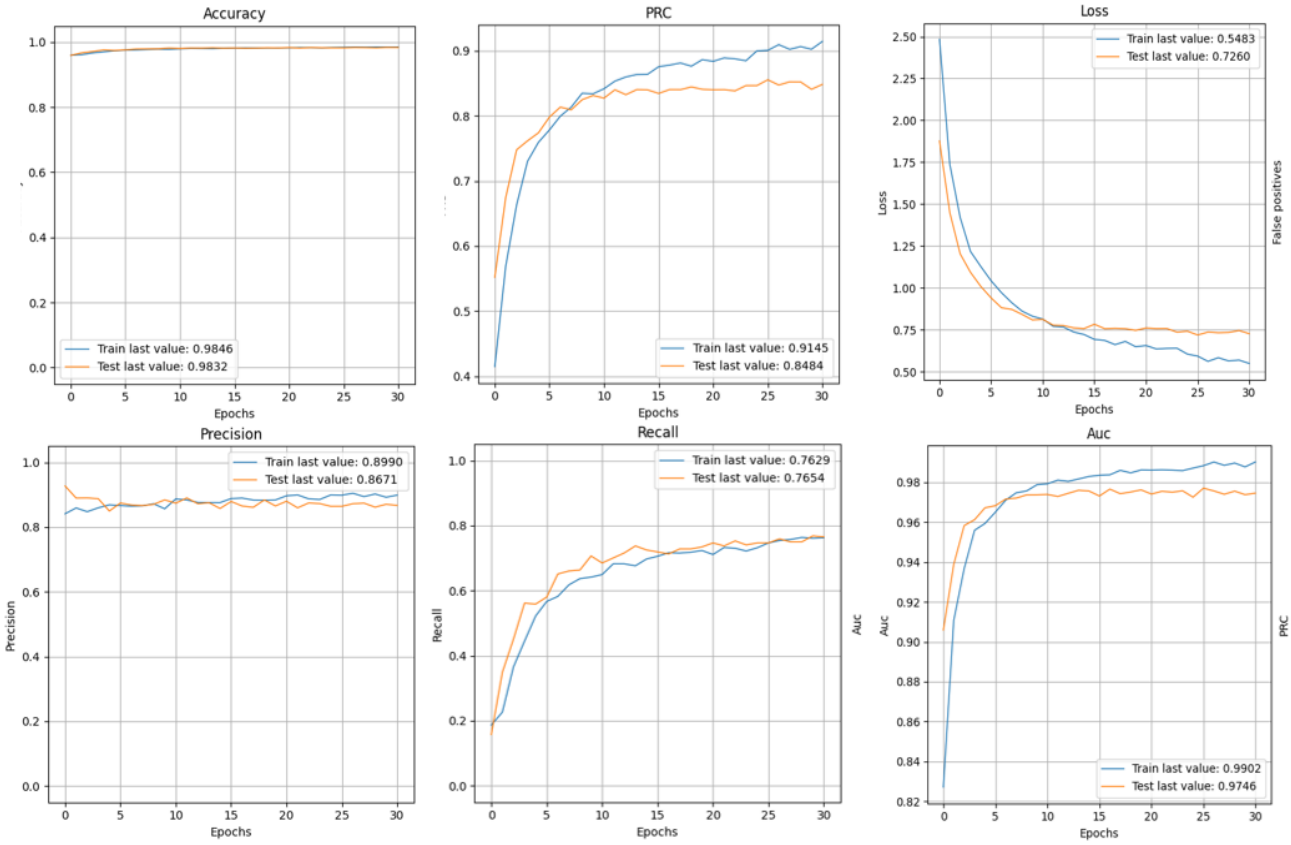


Рис. 4. Графіки метрики навчання нейронної мережі з найкращими результатами

З перших ітерацій модель Adam-[64]-sigmoid-128 продемонструвала хороші результати, що стало можливим завдяки використанню цифрових представлень Telegram-каналів, отриманих на попередньо навченій моделі SBERT, яка спеціалізується на ефективних векторних представленнях речень.

Використання, як входу класифікатора, вектора створеного на основі двох підходів, що включав поєднання інформації з окремих публікацій та статистично виділені ключові та словосполучення, продемонстрував кращі результати ніж кожен з підходів поодинці.

Необхідно відзначити, що отримана перевага має незначний характер, проте вона значуща.

Такий спосіб створення цифрового представлення дозволив узагальнити, зберегти ключові теми та суттєву інформацію з різних публікацій Telegram-каналів в одному компактному представленні.

Висновки

1. Для ембедінгу Telegram-каналів на мережі SBERT визначено такі підходи:

агрегація векторів публікацій,

конкатенація ключових слів за методом TF-IDF,

поєднання перших двох підходів.

Таке рішення дозволило вирішити проблему обмеженості кількості вхідних токенів у моделі SBERT та підвищити ефективність обробки текстової інформації, забезпечуючи можливість в обмеженому за розміром векторі представити більшість ключових особливостей повідомлень Telegram-каналів.

2. Представлено конвеєр обробки текстових даних для цифрового представлення Telegram-каналів.

3. Створено датасет з 9753 цифрових представлень повідомлень Telegram-каналів.

4. На основі ресурсу tgstat.com визначено мітки 38 класів для 5691 з 9753, що склали навчальну вибірку.

5. Експериментально визначено оптимальні за точністю гіперпараметри моделей тематичної класифікації:

- за нейромережевою моделлю з одним прихованим шаром на 64 нейрони, які використовували функцію активації Sigmoid та оптимізатор Adam та забезпечили 98.3% точноності;

- за моделлю машинного навчання Logistic Regression з рівнем регуляризації C = 0.1, яка забезпечила 78.7% точноності.

СПИСОК ЛІТЕРАТУРИ

1. Скринінг українського суспільства протягом повномасштабної війни. Національна рада України з питань телебачення і радіомовлення. URL: https://www.nrada.gov.ua/wp-content/uploads/2022/05/GradusResearch_Report_Suspilne_50K_27042022.pdf .
2. Mikolov, T., Chen, K., Corrado, G., & Dean, J. Efficient estimation of word representations in vector space. 2013. *arXiv preprint arXiv:1301.3781*.
3. Pennington, J., Socher, R., & Manning, C. D. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* 2014, October. (pp. 1532-1543).
4. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, *5*, 135-146. 2017.
5. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. 2018.
6. Reimers, N., & Gurevych, I.. Sentence-BERT: Sentence embeddings using siamese BERT-networks. *arXiv preprint arXiv:1908.10084*. 2019
7. Barbaro, F., & Skumanich, A. Addressing socially destructive disinformation on the web with advanced AI tools: Russia as a case study. In *Companion Proceedings of the ACM Web Conference 2023* (pp. 204-207). 2023, April.
8. Wei, F., & Nguyen, U. T. Twitter Bot Detection Using Neural Networks and Linguistic Embeddings. *IEEE Open Journal of the Computer Society*. 2023.
9. Hugging Face – The AI community building the future. Hugging Face. URL: https://huggingface.co/ (date of access: 30.11.2023).
10. NLTK : Natural Language Toolkit. NLTK :: Natural Language Toolkit. URL: https://www.nltk.org/ (date of access: 30.11.2023).
11. Ukrainian-Stopwords. GitHub. URL: https://github.com/skupriienko/Ukrainian-Stopwords (date of access: 30.11.2023).
12. stopwords-iso/stopwords-ru. GitHub. URL: https://github.com/stopwords-iso/stopwords-ru (date of access: 30.11.2023).
13. Korobov M.: Morphological Analyzer and Generator for Russian and Ukrainian Languages // Analysis of Images, Social Networks and Texts,pp 320-332. 2015.
14. scikit-learn: machine learning in Python. scikit-learn. URL: https://scikit-learn.org/ (date of access: 30.11.2023).
15. The Pushshift Telegram Dataset / B. Jason et al. Zenodo. URL: https://zenodo.org/records/3607497 (date of access: 30.11.2023).
16. Dataset-for-teenagers-chat-in-Telegram-groups: Dataset for teenagers' chat in Telegram groups (Persian). GitHub. URL: https://github.com/imRezaAlie/Dataset-for-teenagers-chat-in-Telegram-groups (date of access: 30.11.2023).
17. Temnikova I. TRACES Bulgarian Telegram Dataset Annotated with Linguistic Markers of Lies. Zenodo. URL: https://zenodo.org/records/7614294 (date of access: 30.11.2023).
18. Crypto telegram groups. Kaggle: Your Machine Learning and Data Science Community. URL: https://www.kaggle.com/datasets/aagghh/crypto-telegram-groups (date of access: 30.11.2023).
19. Telegram channels and groups catalog. *TGStat.com*. URL: https://tgstat.com/ (date of access: 30.11.2023).
20. Keras: The high-level API for TensorFlow | TensorFlow Core [Electronic resource] // TensorFlow. – Mode of access: https://www.tensorflow.org/guide/keras (date of access: 08.12.2023)

Received (Надійшла) 09.12.2023
Accepted for publication (Прийнята до друку) 24.01.2024

Digital representations of Telegram channels

S. Shapovalova, A. Sofiienko

Abstract. The subject of research of this article is digital representations of textual information resources on the example of Telegram channels. The purpose of the work is to determine the optimal method of forming digital representations of Telegram channels for further thematic classification. The following tasks are solved in the article:definition of approaches to the formation of the input vector; determination of the stages of text data processing for the digital representation of the Telegram channel; creation of a dataset of digital representations of Telegram channels; dataset marking for solving the classification problem; determination of hyperparameters of optimal classification models. The following results were obtained: a dataset of digital representations of Telegram channels formed on the basis of the SBERT network was created using three approaches: aggregation of publication vectors, concatenation of keywords using the TF-IDF method, and a combination of the first two approaches; it was determined that the approach of concatenation of keywords using the TF-IDF method and the combination of the first two approaches to the formation of digital representations of Telegram channels based on text publications is the most effective for further classification by topic; the optimal hyperparameters of the thematic classification models are determined in terms of accuracy: Logistic Regressio and deep learning neural networks. A promising direction of further research is the evaluation of the application of the proposed digital representations to clustering and search tasks.

Keywords: natural language text processing, BERT, thematic classification of messages, representation learning.

Цивільна безпека

УДК 628.5/8:614.7 **doi: 10.26906/SUNZ.2024.1.165**

Н. Б. Бурдейна

Київський національний університет будівництва і архітектури, Київ, Україна

ДОСЛІДЖЕННЯ РІВНІВ ІНФРАЗВУКУ У НАВЧАЛЬНИХ ПРИМІЩЕННЯХ ТА ВИЗНАЧЕННЯ УМОВ ЇХ НОРМАЛІЗАЦІЇ

Анотація. Дослідження присвячене вирішенню науково-практичної проблеми нормалізації рівнів інфразвуку в загальних навчальних аудиторіях, комп'ютерних класах, спеціалізованих лабораторіях закладів вищої освіти та наданню рекомендації щодо засобів та заходів з їх нормалізації. Проведений аналіз існуючих досліджень, публікацій та прикладних розробок щодо заходів та засобів нормалізації рівнів інфразвуку в навчальних та виробничих умовах. Виконано натурні вимірювання рівнів інфразвуку в приміщеннях закладів вищої освіти. Вимірювання інфразвуку здійснювалося каліброваним приладом ОКТАВА-110А – шумоміром 1 класу, що має вбудовані октавні та третиноктавні фільтри. Виявлено, що в певних приміщеннях закладів вищої освіти значення інфразвуку наближаються до гранично допустимих відповідно до європейських вимог – 90 дБ. Значні різниці показів приладу за шкалами «Lin» та «А» свідчать про суттєву присутність інфразвуку в загальному акустичному забрудненні. При цьому на деяких локаціях навчальних корпусів і прилеглих територій виявлено неочікувану, повторювальну наявність рівнів інфразвуку 95-105 дБ. Ці факти потребують встановлення джерел підвищеного інфразвукового навантаження на акустичне середовище та їх подальших досліджень. Перспективним напрямом підвищення безпеки студентів, викладачів та співробітників є комплексне обстеження приміщень та прилеглої території університету з наступним складанням мапи акустичного забруднення середовища для розробки системи заходів безпеки на принципах розумної достатності.

Ключові слова: інфразвук, навчальні аудиторії, комп'ютерні класи, навчальні лабораторії.

Вступ

Особливостями інфразвуку – механічних пружних хвиль з частотою до 16 Гц, є несприйняття його органами чуття людини, повільне згасання в просторі через велику довжину хвилі, поширення на великі відстані, незначне поглинання елементами конструкцій, будівель і споруд та шумозахисними матеріалами. У той же час інфразвукові хвилі, в залежності від частоти і рівня звукового тиску, здійснюють шкідливий вплив на організм людини. Інфразвук несприятливо діє на нервову, серцево-судинну системи, функціонування процесу дихання, стан слухового та вестибулярного аналізаторів. Людина може відчувати втому, головний біль, запаморочення, нудоту, зниження гостроти слуху і зору, а також відчуття безпричинної паніки та страху.

Для закладів вищої освіти питання дослідження рівнів інфразвуку у навчальних приміщеннях та визначення умов їх нормалізації є особливо важливим, оскільки організм молодих людей знаходиться у стані розвитку і потребує перебування в умовах з низьким рівнем негативних техногенних впливів. В той же час спеціалізовані лабораторії закладів вищої освіти є приміщеннями з певними ризиками щодо професійної безпеки і стану здоров'я студентів, викладачів та співробітників університету. Досліджень щодо фактичних рівнів інфразвуку у навчальних закладах не проводилось, що обумовлює актуальність даного дослідження

Огляд літературних джерел

Аналіз останніх досліджень і публікацій свідчить, що значна кількість зарубіжних та вітчизняних наукових досліджень присвячена питанням джерел випромінювання інфразвуку, закономірностей його поширення і взаємодії в різних середовищах, будівлях і конструкціях, шкідливого впливу на людський організм, у тому числі у виробничих умовах, способів і засобів нормалізації інфразвуку [1–20].

Обмеження впливу інфразвуку у виробничому середовищі в країнах ЄС здійснюється на основі положень Директиви Європейського Парламенту та Ради ЄС 2003/10/ЄС, оцінка ризиків професійного впливу здійснюється відповідно до міжнародного стандарту ISO 1999:2013 «Acoustics. Estimation of Noise Induced Hearing Loss» [21]. Слід зазначити, шо окремого стандарту щодо вимірювань інфразвуку в країнах ЄС та в Україні на даний час не впроваджено. Визначення експозиції інфразвуку в країнах ЄС здійснюється за методичними вимогами ISO 7196:1995 «Acoustics. Frequency-weighting characteristic for infrasound measurements» [22] з частотно зваженою характеристикою «G». В Україні вплив інфразвуку на робочих місцях регламентується «Санітарними нормами виробничого шуму, ультразвуку та інфразвуку» [23]. Для інфразвуку на території житлової забудови в Україні діяли «Санітарні норми допустимих рівнів інфразвуку і низькочастотного шуму на території житлової забудови» [24], які з 01.01.2017 втратили чинність і не мають повноцінної заміни. За цим нормативним документом гранично допустимий рівень звукового тиску в діапазоні октавних смуг із середньо геометричними частотами 2–31,5 Гц дорівнював 90 дБ. В третино октавних смугах рівень звукового тиску приймався 85 дБ.

На сьогодні є чинними методичні вказівки щодо профілактики несприятливого впливу виробничого

інфразвуку на організм підлітків з урахуванням нормативних даних для їх віку та протипоказань для роботи в умовах впливу виробничого низькочастотного звуку та інфразвуку, а також режиму праці та відпочинку [25]. В Україні перебувають на стадії затвердження санітарні норми щодо допустимих рівнів інфразвуку в приміщеннях житлових та громадських будинків та прилеглих територій. У цьому документі пропонують ввести як гранично допустимі еквівалентні рівні інфразвуку у житлових та громадських приміщеннях 88 дБ, на прилеглих територіях 90 дБ за лінійною шкалою.

Мета роботи – вивчення рівнів інфразвуку у загальних навчальних аудиторіях, комп'ютерних класах та спеціалізованих лабораторіях закладів вищої освіти та надання рекомендацій щодо їх нормалізації відповідно до європейської нормативної бази.

Теоретичні основи дослідження

За характером спектра інфразвук поділяють на широкосмуговий і гармонійний, за тимчасовими характеристиками – на постійний і непостійний Постійний інфразвук нормують за рівнем звукового тиску в октавних смугах частот 2, 4, 8 і 16 Гц. Нормованими характеристиками непостійного інфразвуку є еквівалентні за енергією рівні звукового тиску $L_{екв}$, дБ в октавних смугах частот та еквівалентний загальний рівень звукового тиску в дБ, виміряні за шкалою «Lin». Рівні інфразвуку, що коливається в часі та переривчастого інфразвуку, виміряні за шкалою «Lin», не повинні перевищувати 120 дБ. Еквівалентні по енергії рівні у стандартних октавних смугах частот ($L_{екв}$) та еквівалентний загальний ультразвуковий діапазон (в дБ, за шкалою «Lin») визначаються за формулою:

$$L_{екв} = 10\lg\left[\frac{1}{T}\sum_{i=1}^{n} t_i \cdot 10^{0,1L_i}\right],$$

де T – період спостереження, в год; t_i – тривалість дії шуму з рівнем L_i, в год; L_i – логарифмічний рівень інфразвуку в і-й проміжок часу, в дБ; n – загальна кількість проміжків дії інфразвуку.

Для орієнтовної оцінки виразності інфразвуку можна використовувати загальний рівень звуку, виміряний за шкалою «Lin», та експрес-показник Δ – різницю рівнів, виміряних за шкалами «Lin» та «А»:

$$\Delta = L_{Лін} - L_A,$$

Чим більша різниця Δ, тим вагоміший внесок низькочастотних та інфразвукових складових у спектрі досліджуваного шуму. При значеннях показника від 6 до 10 дБ вважається, що є ознаки наявності інфразвуку, при 11–20 дБ – інфразвук помірно виражений; 21–30 дБ – виражений; більше 30 дБ – значний інфразвук.

Методика дослідження

Вимірювання рівнів низькочастотного звуку та інфразвуку здійснювалося каліброваним шумоміром ОКТАВА-110А. Шумомір ОКТАВА-110А призначений для вимірювання звуку, що впливає на людину, на виробництві, у транспорті, в житлових та громадських будівлях тощо. Прилад може використовуватися для вимірювання шумових характеристик машин, вимірювання звукоізоляції, визначення звукової потужності, атестації приміщень. Є шумоміром 1 класу і має вбудовані октавні та третиноктавні фільтри.

Технічні характеристики шумоміра ОКТАВА- 110А:

– діапазон вимірів – 22–139 дБА (з мікрофоном 50 мВ/Па);

– рівень власних електричних шумів – менше 10 дБА;

– фільтр у режимі "Звук" – октавні фільтри 31,5– 16000 Гц, третиноктавні фільтри 25– 20000 Гц;

– фільтр у режимі "Інфразвук" – октавні фільтри 2– 16 Гц, третиноктавні фільтри 1,6–20 Гц.

Рівні звуку вимірюються за шкалою корекції «А». Рівні інфразвуку вимірюються за шкалою «Lin». Обов'язковими є дані щодо еквівалентних рівнів звуку за шкалою «А». Різниця показів більше за 10 дБ свідчить про суттєву наявність інфразвуку. Вимірювання рівнів звуку здійснювались в октавних та третиноктавних смугах частот як для звуку так і для інфразвуку. Це обумовлено різними нормативними вимогами до звуку в окремих смугах.

Результати дослідження

Дослідження виконувалися у ряді профільних лабораторій КНУБА та комп'ютерних класах. Усі навчальні приміщення розташовані у різних частинах навчальних корпусів. Це виключає вплив стороннього джерела інфразвуку на результати вимірювань. Вимірювання здійснювалися в таких спеціалізованих лабораторіях – лабораторії будівельних машин кафедри будівельних машин, навчальній лабораторії машин та обладнання виробництва будівельних матеріалів і конструкцій кафедри машин і обладнання технологічних процесів та в двох навчальних лабораторіях кафедри фізики – механіки; коливальних, хвильових процесів та оптики.

Отримані дані у третиноктавних смугах частот наведені у таблицях 1–12, де ν – частота інфразвукових хвиль, вимірюється в Гц, L – рівень інфразвуку, вимірюється в дБ.

Таблиця 1 – **Рівні інфразвуку в комп'ютерному класі на 9 комп'ютерів в центральній частині головного корпусу на 6-му поверсі**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	99	96	101	99	95	100	94	90	105	88	72	54

Рівень гучності звуку за шкалою корекції «А» становить 58 дБ, рівень інфразвуку за шкалою «Lin» – 92 дБ, різниця показів – 34 дБ, що свідчить про дуже значну наявність інфразвуку.

Таблиця 2 – **Рівні інфразвуку в комп'ютерному класі на 15 комп'ютерів в правому крилі головного корпусу на 3-му поверсі**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	68	69	64	61	62	57	59	68	69	52	64	65

Рівень гучності звуку за шкалою корекції «А» становить 56 дБ, рівень інфразвуку за шкалою «Lin» – 69 дБ, різниця показів – 13 дБ, що свідчить про суттєву наявність інфразвуку.

Таблиця 3 – **Фонові рівні інфразвуку в лабораторії будівельних машин кафедри будівельних машин**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	57	55	57	51	47	44	41	43	43	41	61	54

Рівень гучності звуку за шкалою корекції «А» становить 44 дБ, рівень інфразвуку за шкалою «Lin» – 58 дБ, різниця показів – 14 дБ, що свідчить про суттєву наявність інфразвуку.

Таблиця 4 – **Рівні інфразвуку в лабораторії будівельних машин кафедри будівельних машин на працюючому стенді для вимірювання опору ґрунтів різанню**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	65	63	59	55	52	47	44	43	42	42	61	50

Рівень гучності звуку за шкалою корекції «А» становить 74 дБ, рівень інфразвуку за шкалою «Lin» – 76 дБ, різниця показів – 2 дБ, що свідчить про несуттєву наявність інфразвуку.

Таблиця 5 – **Рівні інфразвуку в лабораторії будівельних машин кафедри будівельних машин при увімкненій кабіні керування моделі баштового крану КБ 403А**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	79	68	60	60	64	57	48	46	71	52	57	41

Рівень гучності звуку за шкалою корекції «А» становить 61 дБ, рівень інфразвуку за шкалою «Lin» – 82 дБ, різниця показів – 21 дБ, що свідчить про суттєву наявність інфразвуку.

Таблиця 6 – **Рівні інфразвуку в лабораторії будівельних машин кафедри будівельних машин при працюючій моделі баштового крану КБ 403А**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	78	67	61	66	59	68	64	42	48	42	52	48

Рівень гучності звуку за шкалою корекції «А» становить 49 дБ, рівень інфразвуку за шкалою «Lin» – 62 дБ, різниця показів – 13 дБ, що свідчить про суттєву наявність інфразвуку.

Таблиця 7 – **Рівні інфразвуку в лабораторії будівельних машин кафедри будівельних машин при працюючому тельфері (кран-балці)**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	68	64	57	49	47	49	42	44	48	53	67	56

Рівень гучності звуку за шкалою корекції «А» становить 67 дБ, рівень інфразвуку за шкалою «Lin» – 82 дБ, різниця показів – 15 дБ, що свідчить про суттєву наявність інфразвуку.

Таблиця 8 – **Рівні інфразвуку в лабораторії будівельних машин кафедри будівельних машин при увімкненому деревообробному верстаті**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	78	69	68	62	67	82	52	48	57	46	58	48

Рівень гучності звуку за шкалою корекції «А» становить 78 дБ, рівень інфразвуку за шкалою «Lin» – 98 дБ, різниця показів – 20 дБ, що свідчить про суттєву наявність інфразвуку.

Таблиця 9 – **Фонові рівні інфразвуку в навчальній лабораторії машин та обладнання виробництва будівельних матеріалів і конструкцій кафедри машин і обладнання технологічних процесів**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	64	60	57	54	49	46	46	39	36	36	44	39

Рівень гучності звуку за шкалою корекції «А» становить 40 дБ, рівень інфразвуку за шкалою «Lin» – 51 дБ, різниця показів – 11 дБ, що свідчить про суттєву наявність інфразвуку.

Таблиця 10 – **Рівні інфразвуку в навчальній лабораторії машин та обладнання виробництва будівельних матеріалів і конструкцій кафедри машин і обладнання технологічних процесів при увімкненому вібромайданчику**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	65	68	64	52	49	46	47	59	39	37	35	35

Рівень гучності звуку за шкалою корекції «А» становить 65 дБ, рівень інфразвуку за шкалою «Lin» – 72 дБ, різниця показів – 7 дБ, що свідчить про незначну наявність інфразвуку.

Таблиця 11 – **Рівні інфразвуку в навчальній лабораторії механіки кафедри фізики**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	78	76	73	63	58	56	53	50	52	50	46	43

Рівень гучності звуку за шкалою корекції «А» становить 70 дБ, рівень інфразвуку за шкалою «Lin» – 74 дБ, різниця показів – 4 дБ, що свідчить про несуттєву наявність інфразвуку.

Таблиця 12 – **Рівні інфразвуку в навчальній лабораторії коливальних, хвильових процесів та оптики кафедри фізики**

ν, Гц	1,6	2,0	2,5	3,15	4,0	5,0	6,3	8,0	10,0	12,5	16,0	20,0
L, дБ	68	65	61	56	51	50	47	46	52	55	48	44

Рівень гучності звуку за шкалою корекції «А» становить 47 дБ, рівень інфразвуку за шкалою

«Lin» – 56 дБ, різниця показів – 9 дБ, що свідчить про несуттєву наявність інфразвуку.

Аналіз отриманих даних свідчить, що у частині навчальних приміщень рівні інфразвуку суттєві. Це випливає з різниці показів приладу за шкалами «Lin» та «A». В усіх обмежених приміщеннях рівні інфразвуку не перевищують гранично допустимого значення 90 дБ, але за наявності більшої кількості джерел та зовнішнього впливу на акустичну обстановку у приміщеннях значення рівнів інфразвуку можуть стати критичними.

У процесі вимірювань у рівних місцях навчальних корпусів періодично реєструвалися значення інфразвуку рівнів 75–105 дБ. Ніякої закономірності його появи не встановлено. Джерелами таких коливань можуть бути незбалансовані вентилятори вентиляційних систем, ліфтове обладнання, випадкові співпадання коливань внаслідок дорожнього руху тощо. Тому цей параметр середовища повинен періодично контролюватися. У разі системної появи інфразвуку рівнів 90 дБ і вище необхідно вживати відповідні заходи захисту студентів, викладачів і співробітників. Найбільш ефективним заходом зниження рівнів інфразвуку є його придушення у джерелі генерації. Це потребує ревізії усіх потенційних джерел, які обладнані обертовими механізмами малих частот. Для приміщень, де виконуються роботи, які потребують високої зосередженості – комп'ютерні класи, проектні майстерні, спеціалізовані лабораторії тощо можливим є застосування резонансних захисних панелей. Загальні засади проектування таких панелей надано у роботі [27].

Перспективним напрямом робіт зі зниження несприятливого впливу інфразвуку на студентів, викладачів і співробітників є комплексне обстеження акустичної обстановки у навчальних корпусах університету та прилеглої території з наступним формуванням мапи рівнів інфразвуку, що дозволить розробити адекватні організаційно-технічні заходи захисту людей від несприятливих впливів.

Висновки та перспективи подальших досліджень

1. Встановлено, що у частині навчальних приміщень значення рівнів інфразвуку наближаються до гранично допустимих значень (90 дБ). При цьому значні різниці показів рівнів приладу за шкалами «Lin» та «A» свідчать про суттєву присутність інфразвуку у загальному акустичному навантаженні на навчальне і виробниче середовище.

2. Виявлено періодичну та непередбачувану наявність у навчальних корпусах університету інфразвукових коливань зі значеннями рівнями – 95–105 дБ. Джерела підвищення інфразвукового фону не встановлені, що потребує подальших досліджень.

3. Перспективним напрямом підвищення безпеки студентів, викладачів та співробітників є комплексне обстеження приміщень та прилеглої території з наступним складанням мапи акустичного забруднення середовища. Це дозволить розробити систему адекватних заходів безпеки на принципах розумної достатності.

Список літератури

1. Pawlas K., Wpływ infradźwięków i hałasu o niskich częstotliwościach na człowieka – Przegląd piśmiennictwa. Podstawy i Metody Oceny Środowiska Pracy. 2009. № 2(60), s. 27–64. URL: https://www.researchgate.net/publication/250916608_Wplyw_infradzwiekow_i_halasu_o_niskich_czestotliwosciach_na_czlowieka_-_przeglad_pismiennictwa (дата звернення: 19.05.2023).
2. Augustyńska D. Wartości graniczne ekspozycji na infradźwięki – przegląd piśmiennictwa. PiMOŚP. № 2(60), 2009. – s. 15. URL: https://www.semanticscholar.org/paper/Warto%C5%9Bci-graniczne-ekspozycji-na-infrad%C5%BAwi%C4%99ki-%E2%80%93-Augusty%C5%84ska/ 7536b19280002332fa1b44b8b94bd4adc6c509d2 (дата звернення: 19.05.2023).
3. Approved Code of Practice for the Management of Noise in the Workplace. Standards New Zealand. Published by the Occupational Safety and Health Service. Department of Labour. Wellington. New Zealand. First Edition: September 1996. Revised: October 2002. 67 p. URL: https://docplayer.net/16928591-Approved-code-of-practice-for-the-management-of-noise-in-the-workplace.html.
4. Wegleitung zu den Verordnungen 3 und 4 zum Arbeitsgesetz. Schweizerische Eidgenossenschaft Confederation (SECO) – Staatssekretariat für Wirtschaft. 2012. URL: https://www.seco.admin.ch/seco/de/home/Publikationen_Dienstleistungen/Publikationen_und_Formulare/Arbeit/Arbeitsbedingungen/Wegleitungen_zum_Arbeitsgesetz/wegleitung-zu-den-verordnungen-3-und-4-zum-arbeitsgesetz.html.
5. Storm R. Health risk due to exposure of low frequency noise. Orebro University. Örebro, Sweden. 2009. URL: http://www.diva-portal.org/smash/ get/diva2:273045/FULLTEXT01.pdf (дата звернення: 20.05.2023).
6. Health Effects of Exposure to Ultrasound and Infrasound. RCE-14, Documents of Health Protection Agency. DEFRA. 2010. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/335014/RCE-14_for_web_with_security.pdf (дата звернення: 20.05.2023).
7. Araújo Alves J., Neto Paiva F., Torres Silva L., Remoaldo P. Low-Frequency Noise and Its Main Effects on Human Health—A Review of the Literature between 2016 and 2019. Appl. Sci. 2020, 10, 5205. https://doi.org/10.3390/app10155205.
8. Myshchenko I., Nazarenko V., Stopa M., Maslakiewicz M. Occupational Exposure to Infrasonic and Low Frequency Noise: Actual Problems of Hygienic Standardization. Український журнал Охорона праці. 2021. 17 (4). PP. 235-244. https://doi.org/ 10.33573/ujoh2021.04.235.
9. Van Kamp I., van den Berg F. Health effects related to wind turbine sound, including low-frequency sound and infrasound. Acoustics Australia/ Australian Acoustical Society. 46(82). 2018. PP. 31-57. https://doi.org/10.1007/ s40857-017-0115-6.
10. Baeza Moyano D., Gonzalez Lezcano R. Effects of infrasound on health: Looking for improvements in housing conditions. International Journal of Occupational Safety and Ergonomics. 2022. 28(2). PP. 809-823. https://doi.org/ 10.1080/10803548.2020.1831787.

11. Swen M., Stefan H., Martin H., Susanne K. Can infrasound from wind turbines affect myocardial contractility? A critical review. Noise and Health. 2022. 24(113), PP. 96-106. URL: https://eref.uni-bayreuth.de/id/eprint/73087/ (дата звернення: 23.05.2023).
12. Ascone L., Kling C., Wieczorek J., Koch C., Kühn S. A longitudinal, randomized experimental pilot study to investigate the effects of airborne infrasound on human mental health, cognition, and brain structure. Scientific reports. 2021. 11(1). PP. 1-9. https://doi.org/10.1038/s41598-021-82203-6.
13. Chaitidis G.D., Marhavilas P.K., Kanakaris V. Potential Effects on Human Safety and Health from Infrasound and Audible Frequencies Generated by Vibrations of Diesel Engines Using Biofuel Blends at the Workplaces of Sustainable Engineering Systems. Sustainability. 2022, 14. P. 7554. https://doi.org/10.3390/su14137554.
14. McKenna M.H., McComas S.L., Danielle Whitlow R., Diaz-Alvarez H., Jordan A. M., Daniel Costley R., Simpson C. P. Remote structural infrasound: Case studies of real-time infrastructure system monitoring. Journal of Infrastructure Systems. 2021. 27(3), 04021021. https://doi.org/10.1061/ (ASCE)IS.1943-555X.0000623
15. Keith S.E., Daigle G.A., Stinson M. R. Wind turbine low frequency and infrasound propagation and sound pressure level calculations at dwellings. The Journal of the Acoustical Society of America. 2018. 144(2). P. 981-996. https://doi.org/10.1121/1.5051331.
16. Müller L., Kropp W., Zachos G., Forssén J. Investigating Low Frequency Sound from Traffic in a Living Room Lab. Fortschritte der Akustik, 2021. 4 p.
17. Sihar I. Numerical modelling of transient low-frequency sound propagation and vibration in buildings. Eindhoven: Eindhoven University of Technology. 2022. 213 p.
18. Veldboom E., van der Werf C., Incedalci Z., van den Berg F. The effect of masking noise on persons suffering from a low frequency sound. Applied Acoustics. 2022. Volume 191. https://doi.org/10.1016/j.apacoust.2022.108681.
19. Glyva V., Kasatkina N., Levchenko L., Tykhenko O., Nazarenko V., Burdeina N., Panova O., Bahrii M., Nikolaiev K., Biruk Y. Determining the dynamics of electromagnetic fields, air ionization, low-frequency sound and their normalization in premises for computer equipment. Eastern-European Journal of Enterprise Technologies, 2022, 3(10-117), pp. 47–55. https://doi.org/10.15587/1729-4061.2022.258939.
20. Бурдейна Н.Б. Актуальні напрями удосконалення державних будівельних норм проектування нових і реконструкції існуючих закладів освіти. Містобудування та територіальне планування. Київ. 2023. Вип. 82. С. 43-52. https://doi.org/10.32347/2076-815x.2023.82.43-52.
21. Environmental noise guidelines for the European region. 2018:160. World Health Organization. URL: https://www.euro.who.int/en/publications/ abstracts/environmental-noise-guidelines-for-the-europeanregion-2018 (дата звернення: 18.05.2023).
22. ISO 7196:1995 «Acoustics. Frequency-weighting characteristic for infrasound measurements». Publication date: 1995-03. Number of pages: 6. URL: https://www.iso.org/standard/13813.html (дата звернення: 18.05.2023).
23. ДСН 3.3.6.037-99 Санітарні норми виробничого шуму, ультразвуку та інфразвуку. Постанова Міністерство охорони здоров'я від 01.12.1999 № 37. URL: https://zakon.rada.gov.ua/rada/show/va037282-99#Text (дата звернення: 18.05.2023).
24. СанПіН 42-128-4948-89 «Санітарні норми допустимих рівнів інфразвуку і низькочастотного шуму на території житлової забудови». Розробник: Головний державний санітарний лікар СРСР. [Скасовано згідно з розпорядженням Кабміну від 20.01.2016 № 94-р.] URL: http://online.budstandart.com/ua/catalog/doc-page?id_doc=66167 (дата звернення: 18.05.2023).
25. МУ 2410-81 «Производственный шум и профілактика его неблагоприятного воздействия на организм подростков», М., 1981 г. 36 с.
26. ACGIH Threshold Limit Values (TLVs®) and Biological Exposure Indices (BEIs®). 2010. 116 p. URL: https://www.acgih.org/science/tlv-bei-guidelines/ (дата звернення: 19.05.2023).
27. V. Glyva, O. Zaporozhets, L. Levchenko, N. Burdeina, V. Nazarenko. Methodological Foundations Protective Structures Development For Shielding Electromagnetic And Acoustic Fields. Strength of Materials and Theory of Structures. 2023. Issue No. 110. PP. 245-255. https://doi.org/10.32347/2410-2547.2023.110.245-255

Received (Надійшла) 11.12.2023
Accepted for publication (Прийнята до друку) 07.02.2024

Investigation of infrasound levels in educational premises and determination of conditions for their normalization

N. Burdeina

Abstract. The study is dedicated to solving the scientific and practical problem of normalizing infrasound levels in general classrooms, computer classrooms, specialized laboratories of higher education institutions and providing recommendations on means and measures for their normalization. An analysis of existing research, publications, and applied developments regarding measures and means of normalizing infrasound levels in educational and industrial settings was carried out. On-site measurements of infrasound levels in the premises of higher education institutions were performed. The infrasound measurement was carried out with a calibrated OKTAVA-110A device - a class 1 sound level meter with built-in octave and third-octave filters. It was found that in certain premises of higher education institutions, the infrasound values are close to the maximum permissible in accordance with European requirements - 90 dB. Significant differences in the readings of the device on the "Lin" and "A" scales indicate the significant presence of infrasound in the general acoustic pollution. At the same time, unexpected, repeated presence of infrasound levels of 95-105 dB was found in some locations of educational buildings and adjacent territories. These facts require establishing the sources of increased infrasound load on the acoustic environment and their further research. A promising direction for improving the safety of students, teachers and employees is a comprehensive survey of the premises and the surrounding territory of the university, followed by the drawing up of a map of the acoustic pollution of the environment for the development of a system of safety measures based on the principles of reasonable sufficiency.

Keywords: infrasound, educational classrooms, computer classes, educational laboratories.

УДК 629.5.015.5-6:331.45 **doi: 10.26906/SUNZ.2024.1.170**

В. А. Глива[1], В. М. Гусєв[2], Я. І. Бірук[1], М. С. Кашлев[1]

[1] Київський національний університет будівництва і архітектури, Київ, Україна
[2] Морський фаховий коледж Херсонської державної морської академії, Херсон, Україна

ЗАСАДИ ЗНИЖЕННЯ РІВНІВ НИЗЬКОЧАСТОТНОГО ЗВУКУ ТА ІНФРАЗВУКУ У ВИРОБНИЧИХ ТА ПОБУТОВИХ УМОВАХ

Анотація. Показано, що традиційні звукопоглинальні матеріали неефективні для зниження рівнів низькочастотного звуку та інфразвуку. Для нормалізації цих факторів потрібне розроблення спеціальних захисних конструкцій. Тому доцільне дослідження можливості розроблення як конкретних засобів захисту працюючих та населення від впливу низькочастотного звуку та інфразвуку у конкретних умовах, так і формування загальних засад забезпечення нормативних значень цих факторів. Розроблено вимоги до захисних конструкцій. Головними з них є: застосування звукоізоляційних конструкцій не повинне змінювати основні робочі параметри пристроїв або технологічних процесів; параметри звукоізоляційних конструкцій повинні забезпечувати достатній рівень захисту людей; технології виготовлення захисних конструкцій повинні бути економічно прийнятними. Надано розрахунковий апарат щодо визначення акустичного навантаження на середовище. Показано особливості і надано розрахунки щодо визначення навантаження на середовище низькочастотного звуку та інфразвуку. Надано функції щодо розрахунку параметрів захисних панелей резонансного типу для зниження рівнів низькочастотного звуку та інфразвуку. Недоліком таких конструкцій є налаштування на одну резонансну частоту. При цьому наголошено, що усі наведені співвідношення є емпіричними. Тому реальні показники щодо резонансної частоти за обраних параметрів конструкції можуть суттєво відрізнятися. Це вимагає забезпечення певної широкосмуговості захисної конструкції. Для забезпечення прийнятного захисту в області низьких та інфразвукових частот доцільна двошарова панель. Друга панель налаштовується на резонансну частоту, відмінну від першої. Ці частоти обираються за результатами натурних вимірювань і повинні мати найбільші амплітуди у визначеному діапазони частот. Внутрішня панель робиться перфорованою. Це знижує добротність коливальної системи та робить панель більш широкосмуговою. Проміжок між панелями доцільно заповнювати шумопоглинальними матеріалами, наприклад гранульованим пінополістиролом. Це забезпечує зниження рівня шуму в усьому звуковому діапазоні. Певними недоліками пропонованих конструкцій є необхідність конструювання для конкретних умов – розмірів приміщення, площі стіни тощо.

Ключові слова: низькочастотний звук, інфразвук, резонанс, захисна панель.

Вступ

Зниження акустичного навантаження на виробниче та побутове середовище є одним з пріоритетних напрямів досліджень і розробок у галузі цивільної безпеки. Складовими такого навантаження є звукові коливання низької частоти та інфразвук. Цим факторам приділяється недостатньо уваги через їх менше сприйняття органами слуху або несприйняття інфразвукових коливань. В той же час вони несприятливо впливають на здоров'я та самопочуття людини, що підтверджено низкою сучасних гігієнічних досліджень [1, 2]. Тому низькочастотний звук та інфразвук потребує нормування за амплітудними значеннями [3]. Втім більшість досліджень констатують наявність проблеми та визначають вплив цих факторів на здоров'я працюючих та населення. Майже відсутні роботи щодо заходів та засобів зниження рівнів низькочастотного звуку та інфразвуку. Певним чином це обумовлене складністю вирішення таких задач. Низькочастотний звук та інфразвук мають низьке просторове згасання і майже не екрануються будівельними матеріалами та шумозахисними конструкціями. Вважається, що найбільш дієвим засобом зниження їх амплітуд значень є ліквідація або зменшення генерації у джерелі. Але механічне обладнання дуже різноманітне, у тому числі і низькочастотні обертальні механізми. До того ж їх балансування з метою уникнення низькочастотної складової складне. Тому доцільно дослідити можливість розроблення як конкретних засобів зниження поширення низькочастотних механічних коливань, так і загальних підходів до зниження амплітудних значень низькочастотного звуку та інфразвуку у виробничих та побутових умовах.

Огляд літературних джерел. Моніторинг рівнів низькочастотного звуку та інфразвуку свідчать, що їх еквівалентні значення поблизу транспортних магістралей досягають 90 – 100 дБ за лінійною шкалою [4, 5].

Аналогічні дані отримані для силових трансформаторів у населених пунктах [6].

Генерацію цих факторів досліджено у лабораторних умовах та визначено модельні закономірності їх поширення [7, 8].

Але практично усі ці роботи рекомендують уникнути генерації низькочастотних коливань. Тільки поодинокі дослідження пропонують конкретні технічні рішення. Так, у роботі [9] пропонується комплексний підхід до екранування електромагнітних полів та широкосмугового шуму багатошаровими покриттями. Але це прийнятно для захисту окремих будівель і вимагає великої площі та товщини конструкції.

Частково засіб мінімізації низькочастотного звуку надано у роботі [10]. Але ці рекомендації мають окремий характер і стосуються приміщень з комп'ютерною технікою.

У дослідженні [11] представлено результати проектування захисних конструкцій резонансного типу.

Такі конструкції проектуються для конкретного приміщення з визначеними частотами звуку найбіль-

ших амплітуд. Тому доцільно розробити загальні засади технічного зниження амплітуд низькочастотного звуку та інфразвуку.

Мета роботи – розроблення загальних засад зниження рівнів низькочастотного звуку та інфразвуку у виробничих та побутових умовах.

Викладення основного матеріалу

Обирання геометричних характеристик і параметрів звукоізолюючих конструкцій повинне здійснюватися з урахуванням певних загальних вимог. Це обумовлюється, у першу чергу, необхідністю забезпечення стабільної роботи механізму, який є джерелом звуку, тобто його основної функції. Вимоги щодо проектування захисної конструкції можна умовно розділити на три групи:

1. Застосування звукоізолюючої конструкції не повинне змінювати основні робочі параметри пристроїв або технологічних процесів. Такі зміни повинні перебувати принаймні у допустимих межах.
2. Параметри звукоізолюючих конструкцій повинні забезпечувати достатні, з точки зору охорони праці і безпеки життєдіяльності, коефіцієнти звукоізоляції.
3. Технології виготовлення звукоізоляційних конструкцій повинні бути прийнятними з точки зору витрат часу і коштів, тобто мати прийнятну вартість порівняно з ефективністю застосування самого механізму.

Кожна з цих вимог узагальнена, а пріоритетність визначається у закономірності від специфіки виробничого процесу, умов експлуатації, кількості людей, задіяних у виробничому процесі тощо.

Наприклад, перша вимога може обумовлювати необхідність оцінки зворотнього впливу звукоізолюючої конструкції на працездатність чутливих механізмів електронних пристроїв через формування вторинного акустичного поля. Також слід враховувати, що наявність технологічних отворів може суттєво знижувати ефективність захисної конструкції. Незамкнені конструкції неефективні в області низьких частот. Найбільш прості у виготовленні конструкції – циліндричні, пласкі, не потребують розроблення спеціальних технологій виготовлення, але повітряні проміжки між механізмом і стінкою захисної конструкції сприяють появі резонансних явищ, що знижує загальну звукоізоляцію.

Головними вихідними даними у процесі проектування звукоізоляції є конструкція, кінематичні динамічні характеристики виробничих процесів, режими роботи, відомості про необхідність постійного або періодичного доступу до механізмів, розташування робочих місць, шумові характеристики та потрібний ступінь звукоізоляції.

Шумовими характеристиками механізмів є рівень звукової потужності L_p та рівень звукового тиску L на поверхні S_p, яка проходить крізь робоче місце або робочу зону.

Вони зв'язані між собою співвідношенням:

$$L_p = L + 10\lg S_p .$$

Головною вимогою для коректного визначення потрібної ефективності звукоізоляції є ідентифікація джерел звуку. Вона здійснюється шляхом розрахунків або вимірюванням. Останнє обов'язково здійснювати у октавних (третинооктавних) смугах частот.

Потрібна величина звукоізоляції R визначається як:

$$R = L - L_д + 5 ,$$

$$R = L_p - L_д - 10\lg S_p + 5 ,$$

де $L_д$ – допустимі рівні шуму.

Додаток 5 дБ вважається припустимою розбіжністю звукоізоляції у різних частинах конструкції.

Застосування звукоізолюючих конструкцій на низьких частотах особливо актуальне через те, що застосування традиційних засобів зниження шуму – екранів та вигородок не є ефективним. Крім того, у цьому діапазоні більшість механізмів випромінює звук на дискретних частотах, пов'язаних з резонансами окремих частин механізму. Це має певну небезпеку: якщо резонансна частота однієї зі стінок захисної конструкції збігається з частотою збудження, то рівні звуку за межами захисної конструкції підвищується. Таке явище визначено за результатами натурних вимірювань рівнів низькочастотного звуку.

Оцінювання ефективності зниження рівнів звуку низьких частот значно спрощується у випадку, коли максимальні розміри механізму і захисної конструкції менші за довжину хвилі у повітрі. У цьому випадку джерело генерує високі рівні звуку за відмінної від нуля об'ємної швидкості.

$$V(t) = \int_S \bar{\upsilon}(r_0, t) d\bar{S} ,$$

де $\bar{\upsilon}$ – вектор коливальної швидкості поверхні S випромінювача у точці r_0, $d\bar{S}$ – вектор елементу поверхні, спрямований по нормалі до неї.

На великих відстанях r від джерела головною складовою звукового поля буде сферично-симетрична хвиля, яку створила б пульсуюча сфера малого радіусу з об'ємною швидкістю $V(t)$. Звуковий тиск p, який створює мале пульсуюче тіло на відстані r.

$$p(r) = -i\rho_n \omega V_0 \frac{e^{-ikr}}{4\pi r} ,$$

де ρ_n – густина повітря, $\omega = 2\pi f$ – циклічна частота, k – хвильове число у повітрі, V_0 – амплітуда об'ємної швидкості.

Ефективність зниження рівня звуку конструкцією:

$$R = 20\lg\left|\frac{P_m}{P_k}\right| ,$$

де P_m та P_k – амплітуди звукового тиску відкритого та ізольованого механізму.

На низьких частотах, у тому числі і інфразвукових, ефективними є резонансні панелі. Це пласкі панелі, позаду яких є повітряний проміжок.

Для підвищення загального шумопоглинання він може бути заповнений покритим поглиначем. Це тонка панель з будь-якого пружного матеріалу, розташована на деякій відстані від поверхні монтажу (стіни).

Панель є чистою масою, а повітряний проміжок забезпечує пружність.

Резонансна частота панелі визначається як:

$$f_p = 0{,}16c\sqrt{\frac{\rho}{md}},$$

де c – швидкість звуку у повітрі, ρ – густина повітря, m – питома поверхнева маса, d – товщина повітряного проміжку.

У випадку дифузного падіння пружної хвилі приблизне значення резонансної частоти:

$$f_p = 85\sqrt{md},$$

Для зниження рівнів інфразвуку доцільно застосовувати мембранні поглиначі енергії. Вони є конструкцією, у якій гнучкий матеріал натягнутий на жорсткий каркас.

Резонансна частота таких конструкцій визначається зі співвідношення:

$$f_p = \frac{1}{2}\sqrt{\frac{F}{\rho t l^2 b}},$$

де F – сила тяжіння мембрани; ρ – густина матеріалу; l, t, b – довжина, товщина та ширина мембрани.

Недоліком такого поглинача є те, що він не може бути універсальним. Його розміри завжди прив'язані до розмірів поверхні приміщення і розраховуються, виходячи з частоти переважної амплітуди.

Тобто, знаючи найбільш критичну частоту, визначають інші параметри конструкції, які можуть бути змінені і забезпечувати поглинання коливань потрібної частоти.

Головними недоліками наведених вище співвідношень є те, що вони емпіричні. Це обумовлює певні розбіжності з експериментом.

При цьому резонансні конструкції ефективні на одній частоті, в той час як критичних частот може бути декілька.

Тому у процесі проектування звукоізолюючої конструкції доцільно передбачити наявність ще однієї панелі, паралельної першій, але налаштованої на іншу резонансну частоту.

Якщо цю панель зробити перфорованою, то добротність коливальної системи знижується, тобто резонансна крива стає ширшою. Це робить панель певним чином широкосмуговою.

Резонансна частота перфорованої панелі розраховується із співвідношення:

$$f_p = \frac{\upsilon}{2\pi}\sqrt{\frac{S}{t_{еф} dh}},$$

де υ – швидкість звуку у повітрі, $t_{еф}$ – ефективна товщина панелі, $t_{еф} = t + 0{,}5\sqrt{\pi S}$, t – товщина матеріалу панелі, S – переріз одного отвору, d – відстань між центрами отворів, h – відстань від панелі до поверхні монтажу.

Якщо проміжок між панелями заповнити шумопоглинальним матеріалом, наприклад, гранульованим пінополістиролом, то забезпечується зниження рівнів шуму в усьому звуковому діапазоні.

Найбільш типовими джерелами акустичного шуму у промисловості є шум електричних машин та механізмів.

Шум електричних машин визначається трьома складовими:

– магнітний шум, який створюється коливаннями статора і ротора під впливом внутрішніх магнітних полів;

– аеродинамічний шум, генерований рухом повітряних потоків всередині машини;

– механічний шум, обумовлений вібрацією деталей та вузлів машин від неврівноваженості ротора, роботи підшипників та щіток.

У тихоходних машинах, які здебільшого генерують низькочастотний звук, переважає магнітний шум.

Наприклад, для електричних машин загальнопромислового призначення потужностями 1–100 кВт для відстані від машини 0,5 м існує наближене співвідношення для визначення рівня шуму L:

$$L = 10\lg N + 20\lg n + l_0.$$

де N – номінальна потужність електричної машини, n – частота обертання.

Ця залежність стосується асинхронних двигунів та машин постійного струму.

Натурні вимірювання свідчать, що у низькочастотній області звукового спектра (до 500 Гц) складають для електродвигунів до 70 дБ з яскраво вираженими переважними частотами, що дозволяє застосовувати резонансні панелі.

Для дизельних двигунів цей параметр досягає 100–120 дБ також з піковими частотами. Ці частоти для проектування резонансних панелей необхідно визначити експериментально у кожному окремому випадку.

Висновки

1. Зниження рівнів звуку та інфразвуку можливе за рахунок застосування спеціальних конструкції резонансного типу. Це обумовлено неефективністю у низькочастотній області спектра звукопоглинальних матеріалів та низьким поглинанням цих частот будівельними та конструкційними матеріалами.

2. Наданий розрахунковий апарат дозволяє попередньо оцінити параметри захисної конструкції, виходячи з частот переважної амплітуди. При цьому слід враховувати, що наведені співвідношення є емпіричними, тому реальні резонансні частоти можуть дещо відрізнятися від розрахункових. Обов'язковим є тестування дослідних зразків, виходячи з реальних виробничих умов.

3. Загальним недоліком резонансних конструкції для зниження рівнів низькочастотного

звуку та інфразвуку є налаштованість на одну переважну частоту. Для забезпечення певної широкосмуговості конструкції доцільно застосовувати дві паралельні панелі, налаштовані на різні частоти. За умови застосування перфорації на одній з панелей добротність коливальної системи значно знижується, що розширює смугу поглинання. Заповнення проміжку між панелями звукопоглинальним матеріалом забезпечує шумопоглинання в усьому звуковому діапазоні.

СПИСОК ЛІТЕРАТУРИ

1. McKenna M.H., McComas S.L., Danielle Whitlow R., Diaz-Alvarez H., Jordan A. M., Daniel Costley R., Simpson C. P. Remote structural infrasound: Case studies of real-time infrastructure system monitoring. Journal of Infrastructure Systems. 2021. 27(3), 04021021. https://doi.org/10.1061/ (ASCE)IS.1943-555X.0000623
2. Keith S.E., Daigle G.A., Stinson M. R. Wind turbine low frequency and infrasound propagation and sound pressure level calculations at dwellings. The Journal of the Acoustical Society of America. 2018. 144(2). P. 981-996. https://doi.org/10.1121/1.5051331.
3. Myshchenko I., Nazarenko V., Stopa M., Maslakiewicz M. OCCUPATIONAL EXPOSURE TO INFRASONIC AND LOW FREQUENCY NOISE: ACTUAL PROBLEMS OF HYGIENIC STANDARDIZATION. Український журнал Охорона праці. 2021. 17 (4). PP. 235-244. https://doi.org/ 10.33573/ujoh2021.04.235.
4. Кузнецова Е.Б., Булавина И.Д. Особенности мониторинга инфразвукового загрязнения селитебных территорий, прилегающих к транспортным магістралям. Гигиена и санитария. 2018. №12. С. 1141-1145.
5. Гагарин С.А., Рожихин Н.С., Романов Л.И. Трамвай как источних низкочастотного звука и инфразвука. Вестник Удмуртского университета. Экологические проблемы и природопользование, т. 25, выпуск 4. 2015. С. 7-13.
6. Зинкин В.Н., Солдатов С.К., Богомолов А.В., Драган С.П. Актуальные проблемы защиты населения от низкочастотного шума и инфразвука. Технологии гражданской безопасности. 2015. Том 12. № 1 (43). С. 90-96.
7. Müller L., Kropp W., Zachos G., Forssén J. Investigating Low Frequency Sound from Traffic in a Living Room Lab. Fortschritte der Akustik, 2021. 4 p.
8. Sihar I. Numerical modelling of transient low-frequency sound propagation and vibration in buildings. Eindhoven: Eindhoven University of Technology. 2022. 213 p.
9. Nazarenko V.I., Leonov Yu.I., Glyva V.A., Burdeina N.B., Cherednichenko I.M., Pochta V.N., Holubeva A.O. The influence of UV-LED lamps radiation on indicators of microflora in university auditoriums. Ukrainian journal of occupational health. 2023. Vol. 19. № 1. P. 42-50. https://doi.org/10.33573/ujoh2023.01.042
10. Glyva V., Kasatkina N., Levchenko L., Tykhenko O., Nazarenko V., Burdeina N., Panova O., Bahrii M., Nikolaiev K., Biruk Y. Determining the dynamics of electromagnetic fields, air ionization, low-frequency sound and their normalization in premises for computer equipment. Eastern-European Journal of Enterprise Technologies, 2022, 3(10-117), pp. 47–55. https://doi.org/10.15587/1729-4061.2022.258939.
11. V. Glyva, O. Zaporozhets, L. Levchenko, N. Burdeina, V. Nazarenko. Methodological Foundations Protective Structures Development For Shielding Electromagnetic And Acoustic Fields. Strength of Materials and Theory of Structures. 2023. Issue No. 110. PP. 245-255. https://doi.org/10.32347/2410-2547.2023.110.245-255

Received (Надійшла) 29.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Principles of reducing the levels of low-frequency sound and infrasound in production and domestic conditions

V. Glyva, V. Gusev, Y. Biruk, M. Kashlev

Abstract. Traditional sound-absorbing materials have been shown to be ineffective in reducing low-frequency sound and infrasound levels. The normalization of these factors requires the development of special protective structures. Therefore, it is appropriate to study the possibility of developing both specific means of protecting workers and the population from the impact of low-frequency sound and infrasound in specific conditions, as well as the formation of general principles for ensuring the normative values of these factors. Requirements for protective structures have been developed. The main ones are: the use of soundproofing structures should not change the basic operating parameters of devices or technological processes; parameters of soundproofing structures must provide a sufficient level of protection for people; manufacturing technologies of protective structures must be economically acceptable. A calculation device for determining the acoustic load on the environment is provided. Features are shown and calculations are provided for determining the load on the medium of low-frequency sound and infrasound. Functions are provided for calculating the parameters of protective panels of the resonance type for reducing the levels of low-frequency sound and infrasound. The disadvantage of such structures is the adjustment to one resonant frequency. At the same time, it is emphasized that all the given ratios are empirical. Therefore, the real indicators of the resonant frequency under the selected design parameters may differ significantly. This requires ensuring a certain bandwidth of the protective structure. To ensure acceptable protection in the area of low and infrasound frequencies, a two-layer panel is advisable. The second panel is tuned to a resonance frequency different from the first. These frequencies are selected based on the results of field measurements and should have the largest amplitudes in the specified frequency range. The inner panel is made perforated. This reduces the Q-factor of the oscillating system and makes the panel more broadband. It is advisable to fill the space between the panels with noise-absorbing materials, for example, granulated polystyrene. This provides a reduction in the noise level in the entire sound range. Certain disadvantages of the proposed designs are the need to design for specific conditions – room dimensions, wall area, etc.

Keywords: low-frequency sound, infrasound, resonance, protective panel.

УДК 331.453:613.155 (043.5) **doi: 10.26906/SUNZ.2024.1.174**

В. А. Глива[1], О. М. Тихенко[2], Г. Ю. Краснянський[1], С. В. Зозуля[2]

[1] Київський національний університет будівництва і архітектури, Київ, Україна
[2] Національний авіаційний університет, Київ, Україна

ДОСЛІДЖЕННЯ ДИНАМІКИ КОНЦЕНТРАЦІЙ АТМОСФЕРНИХ АЕРОЗОЛІВ, ПИЛУ ТА АЕРОІОНІВ

Анотація. Проведено дослідження змін концентрації атмосферних аероіонів та завислих частинок. Дослідження здійснювалося за припущення, що легкі аероіони однаковим чином осідають на частинки аерозолів та дрібнодисперсного пилу. За вихідні дані необхідно брати генерацію аероіонів внаслідок природної радіоактивності, яка є у даному випадку головним фактором іонізації. У відповідних рівняннях коефіцієнти рекомбінації аероіонів, осідання на нейтральні та протилежно заряджені завислі частинки бралися середніх значень, які відомі з довідкових джерел. Наявність електричного поля масиву аероіонів не враховувалося через відносно малі концентрації аероіонів у атмосферному повітрі за нормальних умов. Проведено верифікацію результатів розрахунків у тестовому приміщенні з відомою генерацією аероіонів та завислих частинок у вигляді аерозолів. Результати тестування показали прийнятний збіг розрахованих та експериментальних даних. Невідповідності, зокрема немонотонність кривих змін концентрацій аероіонів та аерозолів обумовлена великими паспортними похибками вимірювальної апаратури та впливом аерозолів на вимірювання концентрації аероіонів. Для оцінки динаміки концентрацій аероіонів та завислих частинок у реальній атмосфері слід враховувати градієнт та спрямований рух аероіонів у приземному шарі повітря та неоднозначний вплив відносної вологості повітря на концентрації аероіонів. Враховуючи складність вимірювань малих концентрацій у повітрі частинок усіх категорій, розрахунковий метод можна вважати цілком прийнятним.

Ключові слова: аероіони, аерозоль, пил, завислі частинки, коефіцієнт рекомбінації.

Вступ

Концентрації аерозолів, пилу та аероіонів є важливим показником якості повітря. Тому дослідженню змін концентрацій завислих частинок та аероіонів приділяється багато уваги. Більшість робіт у цьому напрямі стосується великих концентрацій пилу у повітрі промислових майданчиків, промислових підприємств – кар'єрів, гірничо-збагачувальних комбінатів, об'єктів будівництва. Це ж стосується аерозолів. Зміни концентрацій аероіонів визначаються здебільшого у приміщеннях у залежності від радіаційного фону, електризації поверхонь, рекомбінації іонів різної полярності тощо. При цьому недостатньо уваги приділяється дослідженню їх взаємодії. Відомо, що аероіони осідають на завислі частинки і надають їм електричний заряд, тобто створюються важкі аероіони, які також взаємодіють між собою. Тому актуальною задачею є дослідження одночасних змін концентрацій аероіонів та завислих частинок різного походження.

Огляд літературних джерел. Більшість робіт з цієї проблематики стосується аероіонного режиму приміщень [1, 2]. Вони розглядають залежності концентрацій аероіонів від наявності джерел деіонізації. Такими джерелами є засоби обчислювальної техніки. У дослідженні [3] показано, що існує прямий зв'язок між концентраціями аероіонів, вологістю повітря та складом атмосферного повітря – значенням полярності аероіонів. У роботах [4–6] ґрунтовно досліджено зв'язок між концентраціями аероіонів та напруженостями електростатичних полів. У роботах [7, 8] показана залежність концентрацій пилу та аероіонів. Ці роботи базуються на рівняннях класичних досліджень [9, 10]. Але отримані результати суто теоретичні. Тому частина коефіцієнтів у рівняннях балансу аероіонів не представлено. Виходячи з цього доцільно визначити експериментально зміну концентрацій аероіонів та завислих частинок за рахунок їх взаємодії.

Мета роботи – дослідження змін концентрацій завислих частинок у атмосферному повітрі за рахунок взаємодії з аероіонами.

Викладення основного матеріалу

Процеси взаємодії атмосферних аероіонів із завислими частинками можна описати аналітично за умови наявності експериментальних даних щодо фонових значень концентрацій усіх частинок та відповідних коефіцієнтів щодо їх взаємодії. Такими коефіцієнтами є коефіцієнти рекомбінації легких аероіонів протилежних знаків, коефіцієнти осадження аероіонів на завислі частинки, коефіцієнти осадження аероіонів на важкі іони (ними можна вважати заряджені аерозолі та порошинки).

У загальному випадку рівняння балансу аероіонів та завислих частинок мають вигляд:

$$\frac{dn^-}{dt} = q - \alpha n^- n^+ - \beta^- n^- N^+ - \beta_0^- n^- N,$$

$$\frac{dn^\mp}{dt} = q - \alpha n^- n^+ - \beta^+ n^\mp N^- - \beta_0^+ n^+ N,$$

де n^-, n^+ – концентрації негативних та позитивних легких аероіонів, см$^{-3}$;
q – рівень генерації пар легких аероіонів, см$^{-3}$;
α– коефіцієнт рекомбінації легких аероіонів; β^-, β^+ – коефіцієнти осадження негативних та позитивних аероіонів на важкі; N^-, N^+ – концентрації важких негативних та позитивних аероіонів, см$^{-3}$;
β_0^-, β_0^+ – коефіцієнти осадження негативних та позитивних аероіонів на нейтральні частинки;
N – концентрація нейтральних частинок, см$^{-3}$.

У разі наявності електростатичних полів до рівнянь додаються наступні складові:

– до першого рівняння додається:

$$-\frac{\mu^- En^-}{r^2}+\frac{\mu^- En^-_{r+1}}{(r+1)^2},$$

– до другого рівняння додається:

$$-\frac{\mu^+ En^+}{r^2}+\frac{\mu^+ En^+_{r+1}}{(r+1)^2},$$

де μ^+, μ^- – рухливості негативних та позитивних легких аероіонів, см²/В/с; E – напруженість електростатичного поля на відстані 1 см від його джерела, В/см; r – відстань від джерела електростатичного поля до досліджуваної зони, см; n^+_{r+1}, n^-_{r+1} – концентрації позитивних та негативних аероіонів у 1 см³ об'єму, суміжних з досліджувальними з боку джерела електростатичного поля.

У реальних умовах відкритого повітря концентрації аероіонів не перевищують 1000–1200 см$^{-3}$, а відсутність заряджених поверхонь дозволяє нехтувати наявністю електростатичних полів та загальним електростатичним полем аероіонів. Коефіцієнти рекомбінації іонів у кожному окремому випадку може бути визначений зі значення радіоактивного фону. В умовах динамічної рівноваги кількість генерованих іонів дорівнює кількості рекомбінованих. Розрахунок можна здійснити, виходячи з іонізаційного визначення одиниці експозиційної дози іонізуючого випромінювання ($1P \approx 2,6\cdot10^{-4}$ *Кл/кг*, у 1 м³ повітря генерується іонізаційний заряд $3,4\cdot10^{-4}$ *Кл*).

За нормальних умов коефіцієнт рекомбінації

$$\alpha = 1,67\cdot10^{-6}\, см^3/с .$$

З експериментальних даних відомо, що

$$\beta^- = 4\cdot10^{-6}\, см^3/с,\ \beta^+ = 3,97\cdot10^{-6}\, см^3/с,$$

$$\beta_0^- = 1,67\cdot10^{-6}\, см^3/с,\ \beta_0^+ = 1,2\cdot10^{-6}\, см^3/с .$$

Для оцінки адекватності розрахунків було проведено експериментальне дослідження взаємозалежності концентрацій аероіонів та завислих частинок. Виконання такого експерименту у відкритому атмосферному повітрі ускладнюється малими концентраціями аероіонів та аерозолів і великими похибками вимірювальної апаратури. Тому експеримент здійснювався у тестовому приміщенні. Температура повітря складала 24–25 °С, відносна вологість – 90–100 %. Зміни концентрацій досліджувальних частинок наведено на рис. 1. Отриманий результат свідчить про чітку взаємну залежність концентрацій аероіонів та завислих частинок. Оцінка здійснювалася за негативними іонами.

У загальному випадку це не зовсім коректно, але врахування аероіонів обох знаків обумовлює врахування електричного поля, що неможливо здійснити експериментально. Існуючі прилади не реєструють поля сукупності аероіонів, а наявність заряджених поверхонь автоматично веде до появи градієнтів концентрацій частинок. Також слід враховувати, що в умовах реальної атмосфери поза приміщеннями існує природний градієнт концентрацій аероіонів у приземному шарі і залежність концентрацій аероіонів від відносної вологості повітря (рис. 2).

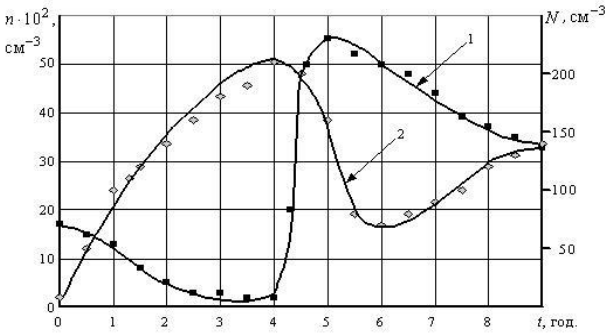


Рис. 1. Зміна концентрацій легких аероіонів (n – графік 1) та завислих частинок (N – графік 2) у приміщенні

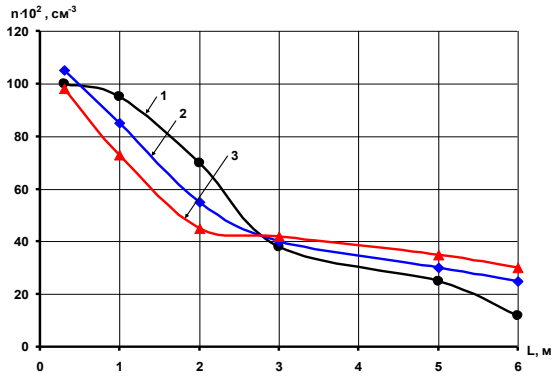


Рис. 2. Зміна концентрацій легких негативних аероіонів з відстанню від джерела для різних відносних вологостей повітря: 1, 2, 3 відповідають відносним вологостям повітря відповідно 40 %, 50 %, 60 %

У даному випадку відстань від джерела іонізації відображає інтенсивність осідання аероіонів на аерозольні частинки. Як видно з рис. 2, за більш високої вологості за великих концентрацій аероіонів вони інтенсивніше осідають на частинки води, утворюючи важкі іони, але зі зменшенням концентрацій ці процеси нівелюються.

Отримані дані взаємної залежності концентрацій аерозолів та аероіонів можуть бути використані для очищення повітря у приміщеннях з великими концентраціями завислих частинок. Як зазначалося вище, аероіони однаковим чином осідають на частинки аерозолів та дрібнодисперсного пилу. Якщо за допомогою іонізатора повітря підвищити концентрації аероіонів, то вони будуть заряджати завислі частинки. В умовах формування електростатичного поля вони будуть дрейфувати у бік поверхонь та осідати на них. Таким чином буде здійснюватися очищення повітря від механічних домішок. Застосування ультразвукового іонізатора дозволить вирішити ще одну задачу. Принцип роботи ультразвукового іонізатора полягає у наступному: коливання ультразвукової частоти подрібнюють воду, яка подається на випромінювач. Внаслідок балоелектричного ефекту відбувається утворення аероіонів. Таким чином відбувається як іонізація повітря, так і його зволоження. Більш вологе повітря сприяє додатковому видаленню дрібнодисперсного пилу.

Представлені математичні співвідношення та реальні коефіцієнти рекомбінації та осідання аероіонів на нейтральні та заряджені завислі частинки з різними зарядами дозволяє попередньо оцінити необхідні продуктивності іонізатора для очищення повітря та ступені деіонізації повітря у залежності від концентрації завислих частинок. У зв'язку цим доцільно доопрацювати наведені співвідношення з метою зниження похибки розрахунків. Для цього необхідно отримати експериментальні дані щодо напруженості електростатичних полів усієї сукупності аероіонів, а також дані щодо напруженості електростатичних полів поверхневих зарядів і пов'язаних з ними градієнтів концентрацій аероіонів поблизу поверхні. Також доцільно визначити переважні знаки поверхневих електростатичних зарядів на типовому обладнанні з полімерними облицюваннями.

В цілому, враховуючи складність одночасних вимірювань концентрацій аероіонів та завислих частинок, розрахунковий метод є прийнятним, для оцінки стану атмосферного повітря.

Висновки

1. Для оцінки стану атмосферного повітря за вмістом аероіонів та зважених частинок доцільно використовувати розрахунковий метод. Для коректного визначення шуканих параметрів необхідно мати надійні дані щодо генерації легких аероіонів. Це можливо здійснювати за значеннями радіоактивного фону.
2. Верифікація результатів розрахунків у тестовому приміщенні за фіксованої генерації аероіонів та аерозолів довела прийнятний збіг розрахункових та експериментальних даних. Велика похибка вимірювань обумовлена впливом завислих частинок на лічильника аероіонів.
3. В умовах реальної атмосфери необхідно враховувати градієнт та спрямований рух аероіонів у приземному шарі повітря та суттєву залежність концентрацій аероіонів та їх просторових змін від відносної вологості повітря.

СПИСОК ЛІТЕРАТУРИ

1. Панова О.В., Левченко Л.О., Теслицький І.А. Дослідження аероіонізації повітря у приміщеннях з експлуатації комп'ютерної техніки. *Комунальне господарство міст*. Сер/: Техн. науки та архітектура. 2021. Т/ 4, № 164. С.215-219.
2. Касаткіна Н.В., Панова О.В., Ніколаєв К.Д. Інноваційні підходи до нормалізації якості повітря виробничого середовища. Збірник наукових праць «Системи управління навігації та зв'язку». Полтава. 2021. Вип. №4 (66) С. 87-89.
3. Bolibrukh, B., Glyva, V., Kasatkina, N., Levchenko, L., Tykhenko, O., Panova, O., Bogatov, O., Petrunok, T., Aznaurian, I., & Zozulya, S. (2022). Monitoring and management ion concentrations in the air of industrial and public premises. Eastern-European Journal of Enterprise Technologies, 1(10(115), 24–30. https://doi.org/10.15587/1729-4061.2022.253110.
4. Глива В. А., Бурдейна Н. Б., Зозуля С. В. Дослідження динаміки аероіон-ного складу повітря на робочому місці користувача персонального комп'ютера з урахуванням електромагнітних чинників. Системи управління, навігації та зв'язку. 2022. Т. 2 (68). С. 99–101
5. Фролов В. Ф., Панова О. В., Зозуля С. В. Прогнозування аероінного складу повітря за наявності природних і штучних джерел іонізації. Комунальне господарство міст. 2022. 1 (168), С. 129–133
6. Глива В. А., Тихенко О. М., Зозуля С. В., Козлітін О. О. Дослідження впливу електростатичних полів на концентрації аероіонів на комп'ютеризо-ваних робочих місцях. Системи управління, навігації та зв'язку. 2023. Т. 2 (72). С. 179–182.
7. Водяник А.О., Бесараб О.М., Сербінова Л.А. Методи й засоби визначення аероінного складу повітря та концентрацій пилу в виробничих умовах. Проблеми охорони праці в Україні. 2011. № 20. С. 66–70.
8. Сукач С.В., Сидоров О.В. Методологічні засади підвищення якості контролю аероіонного складу повітря виробничого середовища. Проблеми охорони праці в Україні. 2016. № 32. С. 127–133.
9. Виснапуу Л.Ю. Электрическое заряжение частиц аэрозоля с применением коронного разряда. Ученые записки Тартуского государственного университета. 1981. Тарту. Вып. 588. С. 77 – 83.
10. Noakes C.J., Sleigh P.A., Beggs C.B. Modelling the air cleaning performance of negative air ionisers in ventilated rooms. Proceeding of the 10 th Int.Conference on Air Distribution in Rooms (Roomvent 2007), 13 – 15 June 2007. – Helsinki, 2007. – 11 p.

Received (Надійшла) 12.12.2023
Accepted for publication (Прийнята до друку) 07.02.2024

Study of the dynamics of atmospheric aerosols, dust and aeroions concentrations

V. Glyva, O. Tykhenko, G. Krasnianskyi, S. Zozulya

Abstract. A study of changes in the concentration of atmospheric aeroions and suspended particles was carried out. The research was carried out under the assumption that light aeroions settle on aerosol particles and fine dust particles in the same way. As the initial data, it is necessary to take the generation of aeroions as a result of natural radioactivity, which is the main factor of ionization in this case. In the corresponding equations, the coefficients of recombination of aeroions, settling on neutral and oppositely charged suspended particles were taken as average values, which are known from reference sources. The presence of the electric field of the array of aeroions was not taken into account due to the relatively small concentrations of aeroions in atmospheric air under normal conditions. Verification of the calculation results was carried out in a test room with a known generation of aeroions and suspended particles in the form of aerosols. The test results showed an acceptable match between the calculated and experimental data. Discrepancies, in particular non-monotony of the curves of changes in the concentrations of aero-ions and aerosols, are caused by large passport errors of the measuring equipment and the influence of aerosols on the measurement of the concentration of aero-ions. To assess the dynamics of concentrations of air ions and suspended particles in the real atmosphere, one should take into account the gradient and directional movement of air ions in the surface layer of the air and the ambiguous influence of relative air humidity on the concentration of air ions. Considering the complexity of measuring small concentrations of particles in the air of all categories, the calculation method can be considered quite acceptable.

Keywords: aeroions, aerosol, dust, suspended particles, recombination coefficient.

УДК 621.384.2.002 **doi: 10.26906/SUNZ.2024.1.177**

Л. А. Зозуля

Національний авіаційний університет, Київ, Україна

ЗАСАДИ РОЗРОБЛЕННЯ БЕЗСВИНЦЕВИХ МАТЕРІАЛІВ ДЛЯ ЕКРАНУВАННЯ ІОНІЗУЮЧИХ ТА НЕІОНІЗУЮЧИХ ЕЛЕКТРОМАГНІТНИХ ВИПРОМІНЮВАНЬ

Анотація. Сучасною тенденцією у галузі захисту населення та працюючих від впливу іонізуючих та неіонізуючих електромагнітних випромінювань є розроблення захисних матеріалів без вмісту свинцю. Існуючі безсвинцеві захисні матеріали мають велику вартість й неприйнятні для масового застосування і облицювання поверхонь великих площ. Розглянуто можливість екранування випромінювань композиційними матеріалами, виготовленими з латексу (матриця) й магнетиту (наповнювач). Доцільність такого підходу обумовлена поширеністю латексів (у тому числі й у рідкому стані) та великим вмістом магнетиту у залізорудному концентраті (більше 80%), який у великих обсягах виробляється гірничо-збагачувальними комбінатами. Теоретично показано, що масові коефіцієнти ослаблення залізовмісних матеріалів, принаймні для випромінювань малих енергій, некритично відрізняються від свинцю. Проведені випробування композиційного матеріалу на основі латексу із вмістом магнетиту з 60% (за масою). Джерелом випромінювання був радіоактивний ізотоп кобальту, який використовується у медичній апаратурі. Результати свідчать, що коефіцієнти екранування (кратності зниження інтенсивності випромінювання) за товщини композиту 1-5 мм складають 1,2 - 3,2; відповідний показник для свинцю тих же товщин – 1,5 – 3,9. Такий результат можна вважати цілком прийнятним з огляду на низьку вартість та екологічність композиту. Перевагою композиту є висока ефективність матеріалу у діапазонах ультрависоких, надвисоких і надзвичайно високих частотах електромагнітних випромінювань. Наявність феромагнітних властивостей у магнетиту забезпечує високі коефіцієнти екранування змінних магнітних полів промислової частоти та стаціонарних магнітних полів діагностичної апаратури. Показана доцільність визначення можливості застосування продуктів очищення промислових стічних вод (сполук важких металів) у якості екрануючих наповнювачів. Потребує досліджень можливість деградації полімерної матриці під впливом іонізуючих випромінювань.

Ключові слова: іонізуюче випромінювання, екранування, композиційний матеріал.

Вступ

Іонізуючі та неіонізуючі електромагнітні випромінювання надвисоких та надзвичайно високих частот шкідливі для людського організму, що підтверджено багатьма медико-гігієнічними дослідженнями [1, 2]. В той же час робочі частоти обладнання бездротового зв'язку постійно підвищуються, що поступово збільшує електромагнітне навантаження на побутове та виробниче середовище. Зростає кількість медико-діагностичного обладнання, яке використовує електромагнітні випромінювання високих енергій. На сьогоднішній день актуальною є задача захисту медичного персоналу від електромагнітних впливів, але існуючі захисні конструкції виготовлені на основі свинцю, який сам по собі є вкрай токсичним, а переробка виробів зі свинцю складна і дороговартісна. Сучасною тенденцією у розробленні матеріалів для захисту від іонізуючих випромінювань є застосування безсвинцевих технологій. Але більшість з таких матеріалів мають велику вартість або масогабаритні параметри, обумовлені малі погонні значення коефіцієнта екранування.

Не дивлячись на панування концепції безпорогового впливу іонізуючих випромінювань Всесвітня організація охорони здоров'я щодо впливу цих та неіонізуючих електромагнітних випромінювань дотримується принципу ALARA (As Low As Reasonably Achievable - настільки низький, наскільки це розумно досяжно).

Тобто у процесі проектування матеріалів та конструкції з них враховуються не тільки технічні можливості, а й співвідношення між витратами та ефектом. Тому актуальною задачею є розроблення матеріалів достатньої ефективності у діапазонах іонізуючих та неіонізуючих електромагнітних випромінювань, виготовлених з компонентів низької вартості з використанням відносно простих технологій.

Сучасний стан питання

Зазвичай матеріали для екранування іонізуючих та неіонізуючих випромінювань розробляються окремо для кожного типу випромінювань і в залежності від діапазону електромагнітних хвиль. Це обумовлено різними фізичними механізмами розсіювання фотонів різної частоти.

У дослідженні [3] представлені результати розроблення матеріалу для екранування гамма-випромінювання на основі вольфраму. Для цього використовувався дрібнодисперсний вольфрам, що автоматично робить його дорогим у виготовленні. При цьому такий матеріал неефективний для поглинання електромагнітних випромінювань неіонізуючих енергій. Частково ці недоліки подолані у матеріалах, описаних у роботах [4, 5]. Певна універсальність досягається за рахунок вмісту заліза у матриці. Але для залізовмісної сполуки у першому випадку необхідна дисперсність майже нанорозмірів, а для другого матеріалу необхідно спеціально виготовляти хромат заліза. До суміші додається октоат кобальту.

Технологія поліефірних композитів складна. Наприклад, для прискорення реакції полімеризації у вихідні компоненти додають пероксид метилетилкетону. Це також ускладнює технологію й підвищує вартість кінцевого матеріалу. Аналогічний недолік притаманний композиційному матеріалу представленому у роботі [6].

Складна сполука заліза, селену та телуру заздалегідь робить матеріал неприйнятним для широкого застосування. До того ж наведені коефіцієнти екранування іонізуючих випромінювань обмежують його застосування випромінюваннями малих інтенсивностей.

Певним компромісом є застосування матеріалів зі зв'язаним свинцем [7] та аморфним залізом [8]. Отримання таких наповнювачів дуже складне, особливо аморфного заліза, яке здебільшого використовується для мінімізації розмірів мікрозондів, узгоджувальних трансформаторів тощо.

У дослідженнях [9, 10] започатковано розроблення безсвинцевих захисних матеріалів для екранування іонізуючих та неіонізуючих випромінювань. Однак перший з них ефективний тільки для м'якого рентгенівського випромінювання, а другий має великі масогабаритні параметри. Але концептуальний підхід застосування у якості матриці та наповнювачів стандартних матеріалів, які у великих обсягах виробляються промисловістю уявляється коректним через можливість отримання захисної композиції, придатної для покриття поверхонь великих площ.

Мета роботи – розроблення загальних засад проектування екологічно чистих захисних матеріалів для екранування іонізуючих та неіонізуючих електромагнітних випромінювань.

Виклад основного матеріалу

У процесі розроблення захисного матеріалу та проєктування захисної конструкції необхідно враховувати багато факторів та обмежень. Особливо це стосується умов, в яких необхідний одночасний захист від іонізуючого та неіонізуючого електромагнітного випромінювання.

Навіть за умови, що ці завдання окремі, бажана наявність універсального матеріалу прийнятних ефективностей в обох частотах електромагнітного спектра. Крім технічного рішення такий підхід й економічно обґрунтований.

У загальному випадку процес проєктування захисного матеріалу або конструкції з нього можна схематично описати певним алгоритмом (рис. 1).

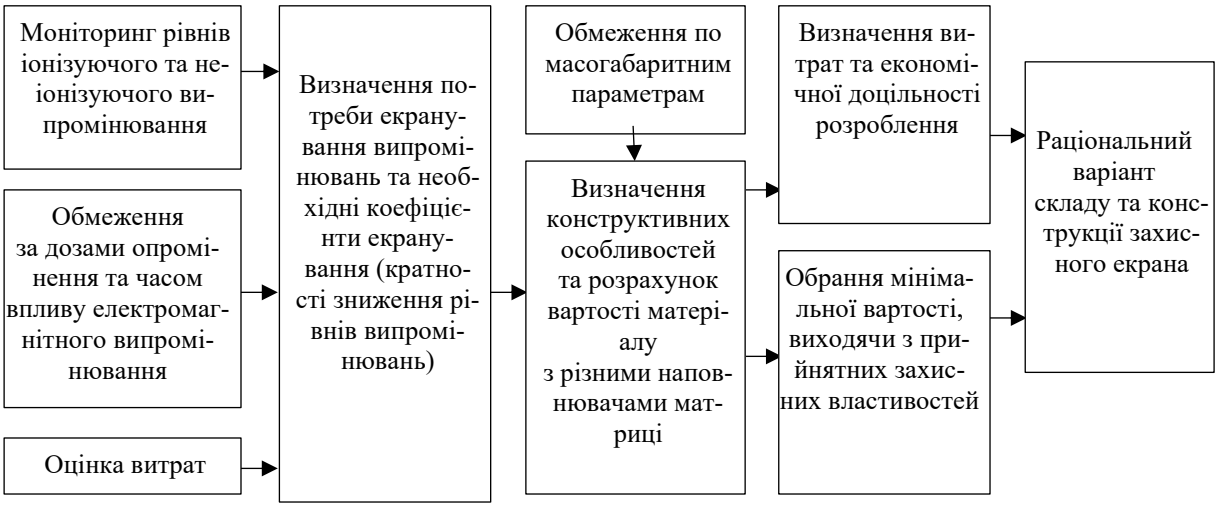


Рис. 1. Процес проєктування захисного матеріалу або конструкції з нього

Для мінімізації витрат коштів та часу на розроблення захисних матеріалів найбільш раціональним є застосування в якості матеріалу матриці компонентів, які у великих обсягах виробляються промисловістю для інших цілей. Це поліетилени, поліпропілени та латекси.

Останні привабливі тим, що виробляються як у вигляді плівок, так і у рідкому стані.

Рідкі латекси швидко полімеризуються за кімнатної температури і добре розчиняють дрібнодисперсні домішки.

В'язкість рідкого латексу дозволяє здійснювати якісне перемішування вихідної суміші ультразвуковим випромінювання. Цей процес добре опрацьований [11] і забезпечує ізотропність суміші, що вкрай важливо для екранування високочастотних електромагнітних випромінювань.

Найбільш критичним фактором щодо екранування іонізуючих випромінювань є матеріал наповнювача матриці, який і визначає ефективність екранування. З метою зниження вартості виробів замість металевого вольфраму можна застосовувати його оксиди, які є вихідними або побічними речовинами основних виробництв.

Достатній ефект, принаймні для енергії випромінювання рентгенівського діагностичного та лікувального обладнання, яке є джерелом гама-випромінювання (побічні складові) дає хром. Розсіювання гама-випромінювання відбувається на атомарному рівні, тому перебування хрому у хімічних сполуках не знижує ефект екранування.

У великих кількостях виробляється окис хрому (тривалентного), який доцільно застосовувати у якості наповнювача.

Найбільш привабливою сполукою для екранування іонізуючих та неіонізуючих електромагнітних випромінювань є магнетит.

Його феромагнітні властивості (відносна магнітна проникність – 4-5) дозволяє ефективно екранувати не тільки електромагнітні високочастотні випромінювання малих інтенсивностей, а й магнітні поля промислової частоти.

Враховуючи значення діелектричних проникностей більшості полімерів, цілком можливо отримати композиційний матеріал з близькими за значеннями магнітної та діелектричної проникностями, що мінімізує коефіцієнти відбиття електромагнітних хвиль. Цей показник є вкрай важливим для захисту людей через унеможливлення перерозподілу випромінювань у приміщеннях через відбиття та перевідбиття електромагнітних випромінювань.

Слід враховувати, що у цьому випадку необхідно чітко визначити товщину матеріалу, достатню для прийнятного рівня поглинання електромагнітної енергії.

Перевагою магнетиту є те, що він є головною складовою залізорудного концентрату, який у великих обсягах виробляється гірничо-збагачувальними комбінатами.

Наприклад, вміст магнетиту у залізорудному концентраті, який виробляється Полтавським гірничо-збагачувальним комбінатом складає принаймні 80-85 %.

При цьому переважною фракцією є частинки дисперсністю до 20 мкм. У стані поставки концентрат схильний до злипання, але ультразвукова обробка суміші дозволяє рівномірно розподілити частинки у полімерній матриці.

Було проведено розрахунки щодо масового коефіцієнту ослаблення гамма-випромінювання композиційного матеріалу з вмістом магнетиту у полімерній матриці (латекс) 45-60 % (за масою).

Розрахунки здійснювалися за припущення, що джерело гама-випромінювання точкове (1 МеВ), детектор має незначні розміри, порівняно з розмірами зразка, зразок має розміри 0,15 х 0,15 м визначеної товщини d.

$$D = D_0 \exp(-\mu d) = D_0 \exp(0{,}693 d/\Delta_{0.5}),$$

де μ – лінійний коефіцієнт ослаблення, d – товщина зразка, $\Delta_{0,5}$ – шар, який знижує інтенсивність випромінювання удвічі, D_0 – потужність дози за $d = 0$.

Масовий коефіцієнт ослаблення μ_m визначається як $\mu_m = \mu/\rho$, де ρ – густина захисного матеріалу.

Встановлено, що масовий коефіцієнт ослаблення матеріалу із вмістом магнетиту складає 0,5–0,6 см²/г. Для свинцю стандартні значення складають 0,07–0,08 (у залежності від енергії випромінювання (1-2 МеВ).

Отриманий результат можна вважати прийнятним, враховуючи, що енергії випромінювань рентгенівських установок масового використання не перевищують кількох десятків кілоелектронвольт.

Було проведено дослідження ефективності екранування матеріалу на основі латексу та магнетиту (60 % за масою) гамма-випромінювання, джерелом якого є ^{60}Co, який застосовується у медичному обладнанні й порівняно результати зі свинцевим екраном (табл. 1, де *Ке* – відношення інтенсивності випромінювання перед екраном до показника у захищеній зоні).

Таблиця 1 – **Ефективність екранування гамма-випромінювання матеріалу на основі магнетиту та свинцю**

Товщина, мм	*Ке*	
	магнетит	свинець
1	1,2	1,5
3	2,1	3,0
5	3,2	3,9

Наведені дані свідчать, що матеріали на основі магнетиту можливо застосовувати для екранування гамма-випромінювань принаймні малих рівнів. При цьому слід враховувати, що випромінювання з боку устаткування, яке використовує джерела іонізуючих випромінювань, ізольовані від зовнішнього простору, тобто вплив на людей складають побічні компоненти випромінювань, які є частково розсіяними й мають малі інтенсивності.

Попередні дослідження довели, що пропонований матеріал дуже ефективний для екранування неіонізуючих електромагнітних полів широкого частотного діапазону, у тому числі й стаціонарних магнітних полів великих напруженостей, які застосовуються у діагностичній апаратурі.

Пропонований напрям розроблення захисних матеріалів уявляється перспективним. Однак потребують подальших досліджень деякі аспекти проблематики.

Відомо, що під впливом навіть ультрафіолетового випромінювання більшість полімерів втрачають вихідні механічні властивості.

Тому необхідно дослідити ступені деградації композиційних матеріалів під впливом іонізуючих випромінювань й з'ясувати їх гарантований ресурс експлуатації.

Доцільно дослідити можливість застосування продуктів очищення промислових стічних вод (наприклад, гальванічного виробництва) для виробництва захисних матеріалів.

Застосування таких відходів дає подвійний ефект (працеохоронний та екологічний) і знижує вартість кінцевої продукції.

Висновки

1. Розроблений алгоритм проєктування матеріалів для екранування іонізуючих та неіонізуючих електромагнітних випромінювань мінімізує час розроблення та дозволяє проєктувати матеріали потрібної ефективності на принципах розумної достатності.

2. Показана можливість застосування для екранування іонізуючих електромагнітних випромінювань магнетиту, який міститься у залізорудному

концентраті, та латексу. Перевагою такого підходу є вироблення цих матеріалів промисловістю у великих обсягах, що здешевлює кінцевий продукт. Порівняння розробленого матеріалу зі стандартними захисними екранами зі свинцю показали прийнятну ефективність композиту принаймні для екранування випромінювань малих енергій.

3. Перспективним напрямом досліджень у цій галузі є визначення можливості застосування у якості екрануючого наповнювача продуктів очищення промислових стічних вод, які мають велику кількість сполук заліза та хрому. Проблемним питанням є деградація полімерної матриці під впливом іонізуючих випромінювань.

СПИСОК ЛІТЕРАТУРИ

1. Duhaini I. The effects of electromagnetic fields on human health. *Physica Medica: European Journal of Medical Physics.* 2016. Vol. 32. P. 213.
2. Seibold, P.; Auvinen, A.; Averbeck, D.; Bourguignon, M.; Hartikainen, J.M.; Hoeschen, C.; Laurent, O.; Noël, G.; Sabatier, L.; Salomaa, S.; et al. Clinical and epidemiological observations on individual radiation sensitivity and susceptibility. *Int. J. Radiat. Biol.* 2019. Vol. 96. P. 324–339.
3. Abu-Al-Roos N. J., Azmana M. N. et al. Tungsten-based material as promising new lead-free gamma radiation shielding material in nuclear medicine. *Physica Medica*. 2020. Volume 78. Pages 48-57. DOI: https://doi.org/10.1016/j.ejmp.2020.08.017.
4. M S Al-Buriahi et al. Fe-based alloys and their shielding properties against directly and indirectly ionizing radiation by using FLUKA simulations. *Physica Scripta.* 2021. Vol. 96. № 4. 045303. DOI: https://doi.org/10.1088/1402-4896/abdd52.
5. Akman F., Ozkan I., Kaçal M.R., Polat H., Issa Shams A.M., Tekin H.O., Agar O. Shielding features, to non-ionizing and ionizing photons, of FeCr-based composites. *Applied Radiation and Isotopes*. Volume 167. 2021. 109470. DOI: https://doi.org/10.1016/j.apradiso.2020.109470.
6. Hamad R.M., Mhareb M.H.A., Alajerami Y.S., Sayyed M.I., Saleh Gameel, Hamad M. Kh, Ziq KhA. A comprehensive ionizing radiation shielding study of $Fe_xSe_{0.5}Te_{0.5}$ alloys with various iron concentrations. *Journal of Alloys and Compounds*. Volume 858. 2021. 157636. DOI: https://doi.org/10.1016/j.jallcom.2020.157636.
7. Rammah Y.S., Olarinoye I.O., El-Agawany F.I., El-Adawy A., Yousef El Sayed. The impact of PbF_2 on the ionizing radiation shielding competence and mechanical properties of TeO_2–PbF_2 glasses and glass-ceramics. *Ceramics International*. Volume 47. Issue 2. 2021. Pages 2547-2556. DOI: https://doi.org/10.1016/j.ceramint.2020.09.100.
8. Alshahrani B., Olarinoye I.O., Mutuwong C., Sriwunkum Chahkrit, Yakout H.A., Tekin H.O., Al-Buriahi M.S. Amorphous alloys with high Fe content for radiation shielding applications. *Radiation Physics and Chemistry*. Volume 183. 2021. 109386. DOI: https://doi.org/10.1016/j.radphyschem.2021.109386.
9. Глива В., Матвєєва І., Левченко Л., Кічата Н. Проектування композитних матеріалів на основі дрібнодисперсної залізовмісної субстанції для екранування іонізуючих випромінювань. *Системи управління, навігації та зв'язку*. 2020. 2 (60). С. 110-113. DOI: https://doi.org/10.26906/SUNZ.2020.2.110.
10. Самченко Д. М., Тихенко О. М., Зозуля Л. А., Цибульник Н. Н. Проектування електромагнітних екранів гарантованої ефективності для галузей цивільної безпеки та електромагнітної сумісності. *Системи управління, навігації та зв'язку.* 2021. 3(73). С. 167−170.
11. Glyva V., Podkopaev S., Levchenko L., Karaieva N., Nikolaiev K., Tykhenko O., Khodakovskyy O., Khalmuradov B. Design and study of protective properties of electromagnetic screens based on iron ore dust. *Eastern-European Journal of Enterprise Technologies.* 2018. Iss. 1/5 (91). P. 10–17.

Received (Надійшла) 29.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Basics of the development of lead-free materials for shielding ionizing and non-ionizing electromagnetic radiation

L. Zozulia

Abstracts. The current trends in the field of protection of the population and workers from the effects of ionizing and non-ionizing electromagnetic radiation are the development of lead-free protective materials. Existing lead-free protective materials are expensive and unacceptable for mass use and for covering large surfaces. The possibility of shielding radiation with composite materials made of latex (matrix) and magnetite (filler) is considered. The expediency of this approach is due to the prevalence of latexes (including in the liquid state) and the high content of magnetite in iron ore concentrate (more than 80%), which is produced in large volumes by mining and processing plants. It has been shown theoretically that the mass attenuation coefficients of iron-containing materials, at least for low-energy radiation, do not differ critically from those of lead. Tests were performed on a latex-based composite material with 60% magnetite content (by weight). The radiation source was a radioactive isotope of cobalt used in medical equipment. The results show that the shielding coefficients (multiples of radiation intensity reduction) for a composite thickness of 1-5 mm are 1.2-3.2; the corresponding figure for lead of the same thickness is 1.5-3.9. This result can be considered quite acceptable given the low cost and environmental friendliness of the composite. The advantage of the composite is the high efficiency of the material in the ranges of ultra-high, ultra-high, and extremely high frequencies of electromagnetic radiation. The presence of ferromagnetic properties in magnetite provides high shielding coefficients for alternating magnetic fields of industrial frequency and stationary magnetic fields of diagnostic equipment. The expediency of determining the possibility of using industrial wastewater treatment products (heavy metal compounds) as shielding fillers is shown. The possibility of degradation of the polymer matrix under the influence of ionizing radiation requires research.

Keywords: ionizing radiation, shielding, composite material.

УДК 628.5:613.6 **doi: 10.26906/SUNZ.2024.1.181**

Д. В. Рєзнік[1], О. О. Ченчева[1], Є. Є. Лашко[1], О. М. Бесараб[2], М. Д. Божик[1]

[1] Кременчуцький національний університет імені Михайла Остроградського, Кременчук, Україна
[2] Київський національний університет будівництва і архітектури, Київ, Україна

ДОСЛІДЖЕННЯ ВПЛИВУ НАГРІВАЛЬНИХ ПРИЛАДІВ І РЕЦИРКУЛЯТОРІВ НА АЕРОІОНИЙ СКЛАД ПОВІТРЯ ВИРОБНИЧОГО ПРИМІЩЕННЯ

Анотація. Мета. Дослідження впливу нагрівальних приладів та рециркулятора на концентрацію та аеродинамічний розподіл іонів обох полярностей у виробничому приміщенні. **Методика.** Дослідження кількісного складу аероіонів здійснювалося за допомогою емпіричного методу під час натурного експерименту, статистичного аналізу та апроксимації отриманих експериментально даних. **Результати.** Ґрунтуючись, на емпіричних даних й отриманих залежностях установлено, що під час використання тепловентилятора збільшується концентрація негативних аероіонів у 1,7 рази, а позитивних у 2,1 рази. Водночас, використання рециркулятора збільшує концентрацію негативних аероіонів у 2,2 рази, а позитивних – 2,8 рази. **Наукова новизна.** Вперше оцінено вплив нагрівальних приладів різного типу та рециркулятора на кількісний склад аерононів у виробничому приміщенні. **Практична цінність.** Установлення закономірностей іонізації (деіонізації) повітря під час використання нагрівальних приладів і рециркулятора у контексті створення безпечного робочого середовища працівників..

Ключові слова: аероіони, мікроклімат, тепловий вентилятор, рециркулятор, інфрачервоний обігрівач.

Вступ

Однією з найважливіших проблем, з якими стикаються сучасні роботодавці, є забезпечення якості повітря виробничого приміщення та забезпечення комфортних мікрокліматичних умов робітників. В холодну пору року часто для досягнення комфортних показників температури використовуються нагрівальні прилади, принцип роботи яких засновано на виділенні променистого або конвекційного тепла. Оцінка впливу такого типу приладів на якісні показники повітряного середовища виробничого приміщення і, як наслідок, на здоров'я і працездатність працівників виробничих приміщень є актуальною задачею в умовах сьогодення. Оскільки людина проводить до 70 % життя у приміщеннях, забезпечення неналежної якості повітряного простору виробничих приміщень може призводити до отруєння продуктами неповного окислення, до дистрофії та атрофії органів і тканин, сприяє передчасному старінню та може стати причиною різних захворювань.

Атмосферне повітря, яким ми дихаємо, несе електричні заряди на частині своїх молекул. Якщо іонізована молекула осіла на частинці рідини або порошинці, такий іон називається важким. Важкі іони шкідливі для здоров'я людини, а легкі, особливо негативно заряджені, мають позитивний вплив на здоров'я людини.

Зовнішнє повітря, проникаючи у виробничі приміщення через вентиляційні установки, втрачає аероіони, особливо легкі з негативним зарядом. Кондиціонування повітря в свою чергу суттєво змінює його електричний стан, а фільтрація через пористі, ватяні, марлеві, масляні та інші фільтри позбавляє повітря всіх аероіонів.

Іонний склад повітря виробничих приміщень набув значної актуальності після того, як було встановлено, що сама людина є джерелом величезної кількості важких іонів (до 500 тис. в 1 $см^3$ повітря). У кожному виробничому приміщенні у присутності людей кількість негативних іонів кисню знижується майже до нуля. Унормування показників аероіонів негативної полярності дозволяє знижувати стомлюваність, втому, відновлювати сили. Все це сприяє покращенню працездатності, посилює імунітет та різко скорочує захворюваність.

Таким чином, метою дослідження є аналіз впливу нагрівальних приладів на аероіонний склад повітряного простори виробничого приміщення.

Аналіз літературних джерел. Велика кількість робіт спрямована саме на дослідження мікрокліматичних показників робочих приміщень його покращенню та дослідженню якості хімічного складу повітряного простору. Основними показниками мікроклімату приміщення, які впливають на стан здоров'я працівників, як правило відносять температуру, вологість, освітлення, якість та швидкість руху повітря, масову концентрацію пилу, електромагнітне випромінювання. Оцінка якості аероіонного складу є важливою задачею, оскільки саме цей показник значною мірою впливає на здоров'я, психічний стан людини, концентрацію уваги під час виконання складних технологічних операцій.

Впливу аероіонів на здоров'я людини присвячено значну кількість наукових праць [2-6].

Результати дослідження

Дослідження проводилось у виробничому приміщенні із застосуванням повірених приладів, а саме: лічильника аероіонів «Сапфір-3К», цифрового термометра ТМ-902CN, цифрового гігрометра Testo 605 H1.

Аналіз аероінного складу повітря при використанні нагрівальних приладів здійснювалося при нагріванні виробничого приміщення загальною площею 50 $м^2$ (загальний об`єм приміщення – 150 $м^3$). Для підвищення температури використали тепловентилятор Expert IFD01-30, рециркулятор РЗТ-300*115 Праймед (Osram) та інфрачервоний обігрівач ECO Mini 1500.

Зовнішній вигляд та паспортні дані нагрівальних приладів представлені у табл. 1.

Таблиця 1 – **Зовнішній вигляд та технічні параметри побутових пристроїв**

Тип пристрою	Зовнішній вигляд	Параметр	Показник
Тепловентилятор Expert IFD01-30		Максимальна потужність	3000Вт
		Нагрівальний елемент	Трубчастий електронагрівач
		Потік повітря	510 м^3/год
		Габарити	305×260×435 мм
		Вага	6,2 кг
Рециркулятор РЗТ-300*115 Праймед (Osram)		Вид	Ультрафіолетова лампа, бактерицидні лампи
		Тип лампи	Безозонова
		Площа приміщення	до 35 м^2
		Потужність	15 Вт
		Живлення	220 В
		Частота	50 Гц
		Тип цоколя лампи	T5
		Виробник лампи	Osram (Німеччина)
		Розміри	2,2×12,4×51,5 см
		Вага	3,5 кг
Інфрачервоний обігрівач ECO Mini 1500		Рекомендована площа відкритого майданчика	5 м^2
		Площа обслуговування	17 м^2
		Нагрівальний елемент	Карбоновий фламентин
		Додаткові характеристики	3 режими обігріву (500/1000/1500 Вт)
		Габарити	23×52,5×14 см
		Вага	2 кг
		Потужність	1500 Вт
		Випромінювання	Короткохвильові (0,78–3 мкм)
		Клас захисту від пилу та вологи	IP24

Під час проведення дослідження у виробничому приміщені вимірювалися мікрокліматичні показники, такі як: температура, вологість повітря та концентрація позитивних та негативних аероіонів. Лічильник іонів було розташовано у геометричному центрі приміщення на висоті 0,9 з метою отримання усереднених даних кількості аероіонів обох полів у виробничому приміщенні. Результати дослідження занесені до табл. 2.

Аналіз отриманих в результаті вимірювання даних, дозволив відмітити наступне:

– прогрівання приміщення з більшою інтенсивністю і швидкістю досягнення бажаних показників здійснюється при використанні інфрачервоного обігрівача. При цьому, повітря у приміщені на 13% більше прогріте ніж при застосуванні теплового вентилятора та на 49 % більше ніж при поступовому прогрівані природнім повітрообміном через вікна;

– показники вологості повітря при використанні інфрачервоного обігрівача знижуються на величину від від 2 до 7% відносно початкового значення зафіксованого у виробничому приміщені. При цьому, застосування теплового вентилятора зменшує цей показник всього на 2 %, інші два прилади надають ще більше зниження;

– концентрація аероінонів (позитивних і негативних) зростає при використанні теплового вентилятора та рециркулятора, а у разі використання інфрачервоного знижується до «нуля».

Аналіз отриманих даних показав, що тепловий вентилятор збільшує концентрацію негативних аероіонів у 1,7 рази, а позитивних у 2,1 рази, що може бути причиною спонуканням до активного руху повітряних шарів у виробничому приміщенні, зумовлених як прискоренням внаслідок руху лопатей вентилятора, так і конвекційними процесами. Застосування рециркулятора дозволяє збільшити концентрацію негативних аероіонів у 2,2 рази, а позитивних – 2,8 рази. Імовірно це може бути викликано додатковим електричним ефектом рециркулятором, що продукує нові аероіони обох полів.

Найсуттєвіший вплив на аероіонний склад повітря виробничого приміщення створює використання інфрачервоного обігрівача. В результаті можна спостерігати ефект, який називається «випалювання повітря».

Отримані результати дозволяють зробити висновок, що з метою швидкого підвищення температури приміщення доцільним є використання інфрачервоною обігрівача в комбінації з рециркулятором, що дозволить стабілізувати та частково компенсувати різке зниження кількості аероіонів у повітряному просторі приміщення.

Найбільш впливовим параметром, від якого залежить концентрація аероіонів, є час роботи пристрою. Зміна показників температури та вологості також залежать від часу, але їх впливом на концентрацію аероіонів можна знехтувати.

Таблиця 2 – **Результати експерименту**

№	Показник				
	Час дослідження, t_1, хв	Відносна температура повітря у приміщені, t_2^*	Відносна вологість повітря у приміщені, φ^*	Концентрація негативних аероіонів, n^-, $см^{-3}$	Концентрація позитивних аероіонів, n^+, $см^{-3}$
Тепловентилятор Expert IFD01-30					
1	0	1,0	1,0	390	760
2	20	1,16	1,0	600	1010
3	40	1,20	1,0	720	1020
4	60	1,25	1,0	730	1100
5	80	1,28	1,0	750	1130
6	100	1,30	0,98	820	1150
7	120	1,35	0,98	860	1220
8	140	1,36	0,98	850	1240
9	160	1,38	0,98	870	1270
10	180	1,39	0,98	860	1290
Рециркулятор РЗТ-300*115 Праймед (Osram)					
1	0	1	1,0	390	760
2	20	1,01	1,0	850	1290
3	40	1,01	0,93	910	1390
4	60	1,01	0,93	1010	1420
5	80	1,01	0,93	1030	1430
6	100	1,02	0,93	1040	1470
7	120	1,02	0,93	1040	1480
8	140	1,03	0,93	1040	1510
9	160	1,03	0,93	1140	1540
10	180	1,03	0,93	1120	1560
	Інфрачервоний обігрівач ECO Mini 1500				
1	0	1,0	1,0	390	760
2	20	1,11	0,98	110	360
3	40	1,24	0,96	0	180
4	60	1,30	0,96	0	90
5	80	1,37	0,96	0	80
6	100	1,41	0,96	0	70
7	120	1,43	0,95	0	40
8	140	1,45	0,95	0	30
9	160	1,49	0,95	0	20
10	180	1,52	0,93	0	10

Примітка * – відносні одиниці, що приведені до нормативних значень

Експериментальні дані було апроксимовано та визначено залежність зміни аероіонного складу повітря, концентрації аероіонів обох зарядів від часу експлуатації нагрівальних приладів. Вираз, який описує дану залежність має вигляд:

$$n^-(t) = a + bt + ct^2 + dt^3 + et^4, \qquad (1)$$

де t – час експлуатації пристрою, хв.; a, b, c, d, e – коефіцієнти апроксимаційного виразу.

Значення коефіцієнтів апроксимаційного виразу наведені у табл. 3.

Експериментальні дані та графіки апроксимованих залежностей концентрації аероіонів від часу роботи апаратів представлено на рис. 1.

Висновки

З метою забезпечення комфортних температурних умов праці працівників виробничих приміщень та задовільного аероіонного складу повітря доцільним є застосування нагрівальних приладів типу теплового вентилятора.

Таблиця 3 – **Коефіцієнти апроксимаційного рівняння**

Тип аероіонів	Коефіцієнти для виразу					Коефіцієнт кореляції
	a	b	c	d	e	
Тепловентилятор Expert IFD01-30						
негативні аероіони	394,69	13,42	-0,19	$-1{,}31*10^{-3}$	$-3{,}19*10^{-6}$	0,96
позитивні аероіони	774,41	12,91	-0,2	$1{,}38*10^{-3}$	$-3{,}35*10^{-6}$	0,95
Рециркулятор РЗТ-300*115 Праймед (Osram)						
негативні аероіони	407,48	26,24	-0,40	$2{,}59*10^{-3}$	$-5{,}77*10^{-6}$	0,95
позитивні аероіони	782,66	31,35	-0,52	$3{,}49*10^{-3}$	$-8{,}1*10^{-6}$	0,96
Інфрачервоний обігрівач ECO Mini 1500						
негативні аероіони	384,69	-18,34	0,29	$-1{,}87*10^{-3}$	$4{,}19*10^{-6}$	0,975
позитивні аероіони	756,43	-25,62	0,35	$-2{,}13*10^{-3}$	$4{,}57*10^{-6}$	0,99

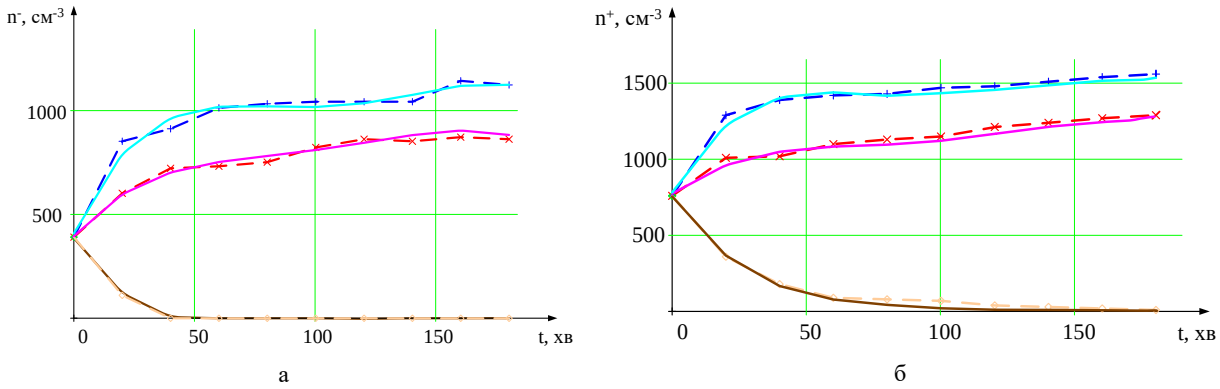


Рис. 1. Динаміка зміни концентрації аероіонів у приміщені: а – негативі аероіони; б – позитивні аероіони – експериментально визначена концентрація аероіонів у разі використання теплового вентилятора Expert IFD01-30; – апроксимована залежність концентрації аероіонів в залежності від часу роботи теплового вентилятора Expert IFD01-30; – експериментально визначена концентрація аероіонів у разі використання рециркулятора РЗТ-300*115 Праймед (Osram); – апроксимована залежність концентрації аероіонів в залежності від часу роботи рециркулятора РЗТ-300*115 Праймед (Osram); – експериментально визначена концентрація аероіонів у разі використання інфрачервоного обігрівача ECO Mini 1500; – апроксимована залежність концентрації аероіонів в залежності від часу роботи інфрачервоного обігрівача ECO Mini 1500

Такі нагрівальні прилади за результатами отриманих експериментальних даних дозволяють збільшити концентрацію негативних аероіонів у 1,7 рази, а позитивних у 2,1 рази, що може бути причиною інтенсивного переміщення повітряних потоків у приміщенні.

Використання рециркуляторів у приміщеннях дозволяє збільшити концентрацію негативних аероіонів у 2,2 рази, а позитивних – 2,8 рази, що імовірно може бути викликано додатковим електричним ефектом, який сприяє виділенню нових аероіонів обох полів. Це позитивно впливає на показники аероіонного складу повітря виробничого приміщення.

При цьому застосування інфрачервоних обігрівачів є небезпечним для нагрівання виробничих приміщень, оскільки призводить до зниження показників аероіонного складу майже до нуля. Використання такого типу обігрівачів вимагає встановлення поруч з ними рециркуляторів з метою компенсації та рекомбінації аероіонів обох зарядів.

СПИСОК ЛІТЕРАТУРИ

1. Shu-Ye Jiang, Ali Ma and Srinivasan Ramachandran Negative Air Ions and Their Effects on Human Health and Air Quality *Improvement. International Journal of Molecular Sciences*. 2018, 19, 2966.
2. Bailey, W.H.; Williams, A.L.; Leonhard, M.J. Exposure of laboratory animals to small air ions: A systematic review of biological and behavioral studies. *BioMed. Eng. online* 2018, 17, 72.
3. Lazzerini, F.T.; Orlando, M.T.; De Prá, W. Progress of negative air ions in health tourism environments applications. *Bol. Soc. Esp. Hidrol. Méd.* 2018, 33, 27–46.
4. Zhou, P.; Yang, Y.; Huang, G.; Lai, A.C.K. Numerical and experimental study on airborne disinfection by negative ions in air duct flow. *Build. Environ.* 2018, 127, 204–210.
5. Ченчева, О. О., Бурдейна, Н. Б., Лашко, Є. Є., Шевченко, В. Г., Петренко І. С. (2022). Вплив пилоутворення при механічному обробленні карбон-карбонових композитів на ризик виникнення професійних захворювань. *Проблеми охорони праці в Україні,* 38(3–4), 25–33.
6. Ченчевой В. В., Сукач С. В., Ченчева О. О., Федорова Н. С., Григор'єва Д. С. Дослідження параметрів гідроаероіонного складу повітря робочого приміщення з ультразвуковою іонізацію. *Вісті Донецького гірничого інституту*. 2020.2(47).168−174.

Received (Надійшла) 11.12.2023
Accepted for publication (Прийнята до друку) 07.02.2024

Research of the influence of heating devices and recirculators on the aeroionic composition of air in the production zone

Dmytro Rieznik, Olga Chencheva, Yevhenii Lashko, Oleg Besarab, Maria Bozhik

Abstract. Pyrpose. Research of the effect of heating devices and a recirculator on the concentration and aerodynamic distribution of ions of both polarities in the production room. **Methodology.** The study of the quantitative composition of aeroions using an empirical method during a field experiment, statistical analysis and approximation of experimentally obtained data. **Findings.** Based on empirical data and obtained dependencies, it was established that during the use of a fan heater, the concentration of negative aerions increases by 1.7 times, and positive ions by 2.1 times. At the same time, the use of a recirculator increases the concentration of negative air ions by 2.2 times, and positive ones by 2.8 times. **Originality.** For the first time, the influence of heating devices of various types and a recirculator on the quantitative composition of aerions in the production zone was evaluated. **Practical value.** Establishing patterns of air ionization (deionization) during the use of heating devices and a recirculator in the context of creating a safe working environment for employees.

Keywords: aerions, microclimate, thermal fan, recirculator, infrared heater.

Зв'язок, телекомунікації та радіотехніка

УДК 621. 39 + 004.057.4 + 519.21 **doi: 10.26906/SUNZ.2024.1.185**

В. М. Воронець, П. Є. Пустовойтов

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

МЕТОД ФОРМУВАННЯ ПЛАНУ ПЕРЕДАЧІ ПАКЕТІВ ПРИ ПІКОВОМУ НАВАНТАЖЕННІ МЕРЕЖІ, ЯКИЙ ЗНИЖУЄ ВІДГУК

Анотація. У представленій статті висвітлено новий підхід до вирішення проблеми оптимізації послідовності передачі групи пакетів у мережі з урахуванням затримки, яка динамічно змінюється у вузлах системи. Автори пропонують використовувати метод мінімаксу для визначення оптимального значення критерію, який моделює динаміку затримки. Важливою частиною методу є декомпозиція вихідної задачі на послідовні двоіндексні підзадачі "про призначення". Ефективність запропонованого підходу підтверджена результатами імітаційного моделювання, що підкреслює його потенційну застосовність у реальних умовах мережевих систем. Показано, що ефективність методу становиться вище, коли кількість пакетів, що передаються, збільшується, та тим більше ефективність, чим більше варіація затримки пакетів у мережі. Таким чином, стаття внесла вагому вклад у розробку методів оптимізації передачі даних у мережах зі змінною динамікою затримки.

Ключові слова: мережний трафік, пікові навантаження, затримка пакета, якість обслуговування.

Актуальність дослідження

Актуальність проблеми управління чергами пакетів у вузлах мережі визначається стрімким ростом обсягу даних та збільшенням завдань, які потребують обробки у реальному часі. Збільшення кількості підключених пристроїв, швидкість передачі даних та розширення функціональності мережевих додатків створюють великий тиск на ефективність управління чергами пакетів. Недостатня пропускна спроможність та неправильна пріоритезація трафіку можуть впливати на якість обслуговування користувачів, спричиняти затримки та погіршувати загальну продуктивність мережі.

Таким чином, вирішення проблем управління чергами пакетів є ключовим аспектом для забезпечення стабільності та оптимальної продуктивності сучасних мереж.

Задача управління чергами пакетів полягає в ефективному розподілі та обробці мережевого трафіку в умовах обмежених ресурсів. Основна мета - запобігти перевантаженню мережевих вузлів, забезпечити справедливий доступ до ресурсів для різних типів трафіку та підтримати найвищий рівень обслуговування для важливих додатків чи послуг. Це включає в себе механізми керування чергами, пріоретизації пакетів, контроль за пропускною здатністю та уникнення затримок, що можуть виникнути в результаті конфліктів в розподілі ресурсів. Задача стає особливо важливою в умовах зростаючого обсягу даних та вимог до якості обслуговування в сучасних мережах.

1. Огляд літератури

Для зменшення затримки пакетів у мережах розглядають аспекти, спрямовані на забезпечення ефективності та якості обслуговування (QoS) [1]. Традиційно у мережах визначають різні класи трафіку та встановлюють пріоритети [2] для обробки, щоб надати важливим додаткам чи послугам перевагу під час передачі даних.

Класифікація трафіку зазвичай відбувається на мережевому рівні моделі OSI (Open Systems Interconnection) [3]. Протокол IP (Internet Protocol), включає поле TOS (Type of Service) чи DSCP (Differentiated Services Code Point), яке використовується для призначення різних класів обслуговування та пріоритетів трафіку в мережі [4].

Хоча поле TOS (Type of Service) байту в заголовку IP-пакету визначає клас обслуговування для пакету, включаючи пріоритети, інші параметри та характеристики, воно само по собі не гарантує вирішення проблем затримок пакетів в мережі. Поле TOS заголовку IP пакету може вказувати на певний клас обслуговування, але він не гарантує конкретних параметрів якості обслуговування, таких як максимальна затримка чи гарантована пропускна здатність [5, 6]. Затримки в мережі можуть виникати через різноманітні причини, такі як переповнення маршрутизаторів, шум на лініях передачі даних або конфлікти ресурсів [7, 8] і поле TOS не може управляти всіма цими аспектами. Також мережі можуть бути дуже неоднорідними, і затримки можуть змінюватися в різних частинах мережі. Поле TOS може надати позначку класу обслуговування, але не завжди гарантує, що цей клас буде однаково врахований на всьому маршруті [9, 10].

Для більш ефективного управління затримками часто використовуються комбіновані підходи, які включають в себе QoS-технології разом із спеціальними механізмами маршрутизації та управління чергами для досягнення кращої продуктивності та надійності мережі.

2. Постановка завдання дослідження

Мета та задачі дослідження. Розробка ефективних методів управління чергами для запобігання переповненню та забезпечення справедливого

доступу до ресурсів для різних користувачів та додатків – одна з найважливіших проблем в мережах електронної комунікації. Для роботи таких методів необхідний постійний моніторинг та аналіз факторів, які впливають на затримки в мережі, для вчасного виявлення проблем та впровадження відповідних коректив, наприклад, вибору ефективних маршрутів для пакетів з метою мінімізації затримок та оптимізації шляхів передачі даних за певними критеріями.

У роботі розглядається спільний критерій мережі – сумарна довжина затримки пакетів, який має велике значення в сфері онлайн управління рухомими об'єктами, оскільки це впливає на реакційний час та загальну ефективність системи.

У зв'язку з цим, **метою статті** є пошук раціональної організації роботи при передачі сукупності пакетів від одного джерела різним адресатам.

В контексті сучасних вимог до мережевого обміну даними та зростаючої потреби в ефективній комунікації, важливо визначити оптимальні методи передачі інформації, що дозволяють забезпечити швидкість, надійність та ефективність обслуговування різних отримувачів. Розглядаються аспекти, пов'язані з оптимізацією передачі пакетів у мережах, з метою вдосконалення загального функціонування систем передачі даних.

Досягнення мети дослідження виконується за рахунок виконання **задач**:

визначення сукупності пакетів та їх характеристик,

визначення вектору порядку передачі пакетів,

оптимізація порядку передачі пакетів,

аналіз результатів оптимізації.

Ці **задачі** спрямовані на створення ефективної та стабільної мережевої інфраструктури, яка задовольняє вимогам сучасних користувачів та застосунків.

Завдання дослідження. Нехай є джерело пакетів, які потрібно передати різним споживачам. Розглянемо задачу відшукання оптимального порядку передачі цих пакетів у припущенні, що кожне з них буде доставлено адресату за оптимальним маршрутом.

Для вирішення задачі передачі сукупності m пакетів будемо виходити с того, що відомий закон $g_k(t)$, $k=1,2,...,K$ зміни у часі довжини черги пакетів на усіх проміжних вузлах мережі. Тоді, використовуючи алгоритм Дейкстри [11, 12] чи Нелдера-Мідда [13] для усіх пакетів, що передаються, можна знайти маршрут, що мінімізує час доставки пакета до одержувача.

Нехай для передачі m пакетів обрано деяку початкову послідовність їх передачі. Така послідовність може бути задана в такий спосіб. Для кожного пакету введемо персональний індикатор x_{ij}, який дорівнює 1, якщо i-й пакет передається j-м за порядком, та дорівнює 0 в іншому випадку.

Тоді введена матриця $X=(x_{ij})$ однозначно задає послідовність передачі пакетів у мережі, якщо для сукупності (x_{ij}), $i=1,2,...,m$, $j=1,2,...,m$, виконуються обмеження:

$$\sum_{i=1}^{m} x_{ij}=1,\ j=1,2,...,m, \tag{1}$$

$$\sum_{j=1}^{m} x_{ij}=1,\ i=1,2,...,m, \tag{2}$$

$$x_{ij}\in\{0;1\},\ i=1,2,...,m,\ j=1,2,...,m. \tag{3}$$

Виконання наведеної сукупності обмежень (1) означає, що на кожне, наприклад, j-е місце в послідовності пакетів, що передаються, призначено для передавання тільки один пакет. Аналогічним чином сукупність обмежень (2) визначає, що кожному пакету з сукупності передачі m призначено тільки одне місце.

Приймемо, що довжини пакетів не дуже відрізняються одна від одної і тривалість передачі для будь-якого з пакетів дорівнює Δ. Тоді під час передачі i-го пакета j-м (тобто для пари (i,j), що має значення $x_{ij}=1$) за порядком знайдемо найкоротший маршрут, який починається в момент $T_j=T_0+(j-1)\Delta$, та відповідний цьому маршруту час затримки доставки T_{ij} пакета цільовому хосту. T_0 – момент початку передачі набору пакетів. Розв'язуючи задачу для всіх пар (i,j), $i=1,2,...,m$, $j=1,2,...,m$, будується матриця $T=(T_{ij})$.

Для вирішення задачі треба розглянути $m!$ різноманітних послідовностей передачі пакетів. Очевидно, що для реальної кількості пакетів m у системі та обмежень на час виконання оптимізації виконувати повний перебір не має перспективи. Тому у роботі пропонується метод пошуку найкращого порядку передачі сукупності пакетів, а критерієм оптимізації пропонується максимальний час доставки пакета. При цьому для обраного плану передачі $X=(x_{ij})$ значення критерію розраховується, як

$$\eta(X)=\max_{i,j}\left\{T_{ij}x_{ij}\right\}. \tag{4}$$

3. Загальні результати

Метод визначення оптимального порядку передачі пакетів полягає у наступному, необхідно знайти план послідовності відправки $X=(x_{ij})$, що мінімізує критерій (4) та задовольняє обмеженням (1) – (3). Отримана задача є мінімаксною задачею «про призначення» [14]. Для її вирішення пропонується така методика.

Упорядкуємо вихідну множину значень T_{ij}, $i=1,2,...,m$, $j=1,2,...,m$ таким чином:

$$T_{i_1j_1}\ge T_{i_2j_2}\ge ...\ge T_{i_qj_q}\ge ...\ge T_{i_{mm}j_{mm}}. \tag{5}$$

Із кожним елементом $T_{i_q j_q}$, $q \in \left\{ 1,2,...,m^2 \right\}$, пов'яжемо двоіндексну матрицю $D^{(q)} = (d_{ij}^{(q)})$, компоненти якої задаються співвідношенням

$$d_{ij}^{(q)} = \begin{cases} T_{ij}, \ якщо\, T_{ij} < T_{i_q j_q}, \\ M, \ якщо\, T_{ij} \geq T_{i_q j_q}, \end{cases} \quad (6)$$

тут M – достатньо велике значення (наприклад, нехай $M = m^2 \max\limits_{ij} \left\{ T_{ij} \right\}$).

Розглянемо випадок $q = 1$. Тут у матриці $D^{(1)} = (d_{ij}^{(1)})$ буде всього один елемент, що дорівнює M, який розташовано на позиції i_1, j_1.

Розв'яжемо тепер задачу пошуку набору $X = (x_{ij})$, який мінімізує

$$L(X) = \sum_{i=1}^{m} \sum_{j=1}^{m} d_{ij}^{(1)} x_{ij} \quad (7)$$

та задовольняє обмеженням (1) – (3).

Така задача «про призначення» традиційно вирішується «угорським» методом [14].

Якщо при цьому значення $L(X_1^*)$ у оптимального плану задачі (1) – (3), (7) менше за M, то це значить, що існує певний порядок передачі пакетів, при якому максимальний час передачі менше за $T_{i_1 j_1}$.

Розглянемо тепер випадок $q = 2$. У відповідній матриці

$$D^{(2)} = (d_{ij}^{(2)})$$

буде вже два елемента, які знаходяться на відповідних місцях (i_1, j_1), (i_2, j_2), та дорівнюють M. Знов використаємо «угорський» метод для розв'язання задачі «про призначення» із відповідною матрицею $D^{(2)}$. Аналогічно попередньому випадку, із виконання нерівності $L(X_2^*) < M$ виходить, що отриманий на цьому кроці порядок передачі пакетів X_2^* забезпечує їх передачу за час, що не перевищує $T_{i_2 j_2}$.

Продовжуючи розв'язання задачі дійдемо до деякого значення $q = \tilde{q}$, такого, що $L(X_{\tilde{q}}) < M$, але $L(X_{\tilde{q}+1}) > M$.

Це означає, що існує такий порядок передачі пакетів, в якому максимальний час передачі не перевищує $T_{i_{\tilde{q}}, j_{\tilde{q}}}$, але не існує порядку, для якого максимальний час менший або дорівнює $T_{i_{\tilde{q}+1}, j_{\tilde{q}+1}}$. Таким чином, отримуємо план $X_{\tilde{q}}^*$ який є розв'язанням вихідної мінімаксної задачі (1) – (4), який шукали.

4. Аналіз результатів дослідження

Проведемо оцінку доцільності використання запропонованого методу оптимізації порядку передачі пакетів. З цією метою здійснимо імітаційне моделювання вузла мережі на певному інтервалі часу його функціонування, він передаватиме m пакетів із різними $g_k(t)$ законами зміни в часі довжини черги пакетів, які очікують на початок обслуговування в проміжних вузлах.

Зрозуміло, що використання раціонального порядку передачі пакетів тим більше доцільно, чим вище варіабельність функції $g_k(t)$, що виявляється в відмінностях величини затримки T_{ij} доставки пакетів, які передаються різними за порядком.

Для оцінювання рівня варіабельності введемо показник

$$\xi = \frac{\max\limits_{i} \{ \max\limits_{j} T_{ij} - \min\limits_{j} T_{ij} \}}{\frac{1}{m^2} \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{m} T_{ij}}. \quad (8)$$

Для оцінки рівня доцільності оптимізації порядку передачі пакетів введемо показник

$$\zeta(X^*) = \frac{\max\limits_{i} T_{ij}}{\max\limits_{ij} T_{ij} x_{ij}^*}. \quad (9)$$

Чисельник цього співвідношення визначає максимальну затримку доставки у випадку, коли пакети передаються відповідно до їх початкових номерів. В займеннику (9) стоїть максимальна затримка доставки, яка відповідає оптимальному порядку передачі пакетів у мережі, що моделюється.

Результати моделювання для різних значень кількості пакетів m, що пересилаються мережею наведено на рис. 1.

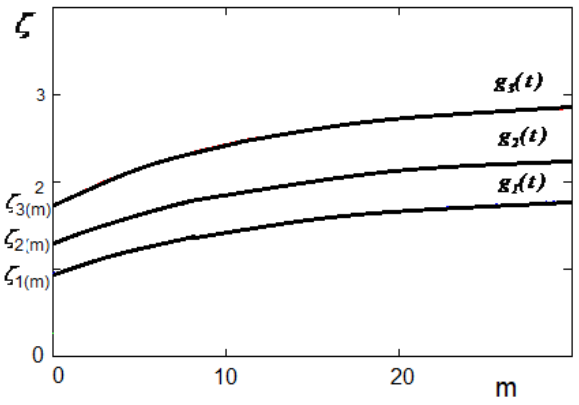


Рис. 1. Залежність рівня доцільності оптимізації ζ від кількості пакетів *m* для різних значень варіації випадкової величини $g_k(t)$

Аналіз наведених кривих дозволяє зробити наступний висновок: виграш ζ, який дає оптимізація порядку передачі пакетів зростає із збільшенням

кількості пакетів m, що передаються, і збільшенням рівня варіабельності довжини черги пакетів $g_k(t)$, що очікують обслуговування в проміжних вузлах.

Висновки

У роботі запропоновано метод розв'язання задачі оптимізації порядку передачі сукупності пакетів з урахуванням динаміки затримки пакетів у вузлах мережі.

При розв'язанні задачі запропоновано критерій, оптимальне значення якого розраховується методом мінімаксу.

Показано, що вихідна задача зводиться декомпозицією до послідовності двоіндексних підзадач «про призначення».

Доцільність оптимізації порядку передачі підтверджена імітаційним моделюванням.

Багато систем управління рухомими об'єктами, такі як автомобільні системи безпеки, дрони, системи телемедицини тощо, вимагають реального часу для надання швидкої та точної реакції на події. Низькі затримки передачі пакетів мережею дозволяють системі сприймати, аналізувати та реагувати на зміни в реальному часі.

У сферах, де важлива точність руху, таких як автономні транспортні засоби або навігація дронів, затримки в передачі даних можуть впливати на точність прогнозування руху та управління маршрутом.

Враховуючи ці фактори, важливо розробляти та впроваджувати методи оптимізації мережевих та комунікаційних технологій, які забезпечують мінімальні затримки для забезпечення ефективного та безпечного управління рухомими об'єктами в режимі реального часу.

СПИСОК ЛІТЕРАТУРИ

1. QoS: Congestion Management Configuration Guide, Cisco IOS XE Everest 16.5, Cisco Systems, Inc., 2019.
2. Пустовойтов П. Е., Раскин Л. Г. "Дисциплина обслуживания в мультисервисных сетях, минимизирующая максимальную задержку." Проблеми телекомунікацій. – 2013. – № 1 (10). – С. 66 – 71.
3. Оптимизация порядка передачи сообщений в узлах компьютерных сетей с учетом динамики трафика / П.Е. Пустовойтов, Л.Г. Раскин // Системні дослідження та інформаційні технології. — 2013. — № 3. — С. 53-57.
4. Пустовойтов, П. Е. Модель узла компьютерной сети с повторной передачей утерянных пакетов / П. Е. Пустовойтов // Прикладная радиоэлектроника: науч.-техн. журн. – Х. : ХНУРЭ, 2012. – Т. 11, № 1. – С. 87–90.
5. Pustovoitov, P., Okhrimenko, M., Voronets, V., & Udalov, D. (2021). The speed calculating increasing method of the markov model network node. *Advanced Information Systems*, *5*(3), 13–17. https://doi.org/10.20998/2522-9052.2021.3.02.
6. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), "Horizontal scaling method for a hyperconverged network", *2021 International Conference on Information and Digital Technologies* (IDT), Zilina, Slovakia, doi: https://doi.org/10.1109/IDT52577.2021.9497534
7. Yang, S., Xu, C., Zhong, L., Shen, J., Muntean, G. M. (2019), "A QoE-Driven Multicast Strategy With Segment Routing–A Novel Multimedia Traffic Engineering Paradigm", IEEE Transactions on Broadcasting, No. 66(1), P. 34-46. DOI: https://doi.org/10.1109/TBC.2019.2932338.
8. Kuchuk, N., Kovalenko, A., Ruban, I., Shyshatskyi, A., Zakovorotnyi, O. And Sheviakov, I. (2023), "Traffic Modeling for the Industrial Internet of NanoThings", *2023 IEEE 4th KhPI Week on Advanced Technology, KhPI Week 2023 - Conference Proceedings*, 194480, doi: http://dx.doi.org/10.1109/KhPIWeek61412.2023.10312856
9. O. Lemeshko and V. Sterin, "Design and structural-functional optimization transport telecommunication network", XIIth International conference the exerience of designing and CAD system (CADSM'2013), pp. 208-210, February 2013.
10. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об'єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: https://doi.org/10.20998/2522-9052.2018.1.04
11. O.V. Lemeshko, S.V. Garkusha, O.S. Yeremenko and A.M. Hailan, "Policy-based QoS management model for multiservice networks", International Siberian Conference on Control and Communications (SIBCON'2013), pp. 1-4, May 2015.
12. Зиков І. С., Кучук Н. Г., Шматков С. І. Синтез архітектури комп'ютерної системи управління транзакціями e-learning. *Сучасні інформаційні системи*. 2018. Т. 2, № 3. С. 60–66. DOI: https://doi.org/10.20998/2522-9052.2018.3.10
13. O. Lemeshko and O. Drobot, "A Mathematical Model of Multipath QoS-based Routing in Multiservice Networks", Modern Problems of Radio Engineering Telecommunications and Computer Science (TCSET'2006), pp. 72-74, February 2006.
14. R.E. Burkard, M. Dell'Amico, S. Martello: *Assignment Problems* (Revised reprint). SIAM, Philadelphia (PA.) 2012.

Received (Надійшла) 25.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

A method for forming a packet transmission plan at peak network load which reduces feedback

V. Voronets, P. Pustovoitov

Abstract. The presented article highlights a new approach to solving the problem of optimizing the sequence of transmission of a group of packets in the network, taking into account the delay that dynamically changes in the nodes of the system. The authors suggest using the minimax method to determine the optimal value of the criterion that models the delay dynamics. An important part of the method is the decomposition of the original problem into successive two-index subproblems "about the destination". The effectiveness of the proposed approach is confirmed by the results of simulation modeling, which emphasizes its potential applicability in real conditions of network systems. It is shown that the efficiency of the method becomes higher when the number of transmitted packets increases, and the greater the efficiency, the greater the variation of packet delay in the network. Thus, the article made a significant contribution to the development of methods for optimizing data transmission in networks with variable delay dynamics.

Keywords: network traffic, peak loads, packet delay, quality of service.

УДК 621.39 **doi: 10.26906/SUNZ.2024.1.189**

В. В. Ганзій, А. А. Коваленко, О. В. Ситник

Харківський національний університет радіоелектроніки, Харків, Україна

АНАЛІЗ МЕТОДІВ УПРАВЛІННЯ ПРОЦЕСАМИ ПЕРЕДАЧІ ДАНИХ ТА ТРАФІКОМ У МУЛЬТИСЕРВІСНИХ КОМП'ЮТЕРНИХ МЕРЕЖАХ

Анотація. Метою даної роботи є проведення аналізу методів управління процесами передачі даних та трафіком які зможуть забезпечити підвищення ефективності передачі даних у мультисервісних комп'ютерних мережах. Зростаючі потреби суспільства в нових послугах телекомунікаційних мереж призводять до зміни ідеології побудови останніх кожне десятиріччя. Сьогодні, на зміну технологіям, що використовують мультиплексування з розділенням та ущільненням за довжиною хвилі, приходять мультисервісні технології, основним принципом концепції яких є відділення одна від одної функцій перенесення та комутації, функцій керування транзакціями та функцій керування послугами. Зокрема, це можуть бути різноманітні корпоративні мережі, де важливо контролювати доступ до ресурсів, хмарні середовищах, де надаються різноманітні послуги, онлайн-ігрові сервіси, де потрібна низька затримка та стабільне з'єднання. Щоб підвищити ефективність мультисервісних комп'ютерних мереж сьогодні інтегрують технічні засоби, що використовують асинхронний режим передачі (технологія ATM або B-ISDN), з застосунками для мережі Інтернет. При побудові таких мереж важливо враховувати класифікацію мережевих характеристик, а саме категорію трафіку для вибору оптимального методу управління. В даній статті розглянуто метод програмного управління параметрами з'єднанням та методи статистичного мультиплексування передачі даних, проаналізовані основні переваги та проблеми означених методів, які вимагають подальшого їх дослідження.

Ключові слова: Метод, управління, процес, мультисервісна комп'ютерна мережа, модель, взаємодія, трафік, застосунок.

Вступ

Ефективність сучасних високошвидкісних комп'ютерних мереж різноманітних класів є критичною характеристикою для багатьох сервісів [1,2]. Для визначе-ння ефективності мультисервісної комп'ютер-ої мережі (МСМ) необхідно дотримуватись основних визначень, які існують у міжнародних стандартах [3]. Базові поняття, на які опираються більшість визначень, є наступними: зв'язок – процес передавання інформації відповідно деяким правилам; з'єднання – деяка асоціація двох або більше пристроїв для здійснення зв'язку між ними; ресурс – загальна назва фізичних або концептуальних сутностей всередині телекомунікаційних мереж, використання яких визначається однозначно; користувач – загальний термін для усіх зовнішніх по відношенню до мереж сутностей, які використовують з'єднання через мережу для комунікації [4–6].

У процесі комунікації користувачів у МСМ виникає потік повідомлень – трафік, який може бути охарактеризований кількісно. Зазвичай для кількісної характеристики трафіка використовується його об'єм. Для цифрових систем ця величина асоціюється з числом бітів, які були передані за заданий час. Однак для аналогових систем такій підхід виявляється неприйнятним. Більш того, при використовуванні цифрових систем зі складними способами модуляції та кодування сигналів, визначення об'єму трафіка стає неоднозначним. Тому кількісною характеристикою об'єму трафіку, який був використаний тим чи іншим ресурсом, визначають величину сумарного, інтегрального інтервалу часу, протягом якого даний ресурс був зайнятий за період часу, що аналізується. Іноді таку величину називають роботою ресурсу протягом заданого часу.

Мета статті – аналіз методів управління процесами передачі даних та трафіком для підвищення ефективності мультисервісної комп'ютерної мережі та виявлення основних переваг і проблем даних методів для подальшого їх дослідження та впровадження.

Аналіз методів управління процесами передачі даних в МСМ

У складному процесі взаємодії інформаційних процесів в комп'ютерних мережах можна виділити ряд аспектів, з яких можна визначити, що методи управління мережевими процесами у віртуальних каналах зв'язку займають одне з центральних місць.

У зв'язку з постійно зростаючими вимогами до швидкості обміну даними розвиток мережевих технологій здійснюється шляхом інтеграції технічних засобів, що використовують асинхронний режим передачі (технологія ATM або B-ISDN), з застосунками для мережі Інтернет. В основі мережевої інтеграції лежать програмно-апаратні засоби передачі даних, що реалізовують управління на різних рівнях моделі міжмережної взаємодії OSI [7,8].

З позицій теорії управління, при розгляді процесів в інтегрованій мережі можна виділити такий важливий аспект, як інформаційні сигнали, що дозволяють здійснювати управління, передаються в одних і тих же фізичних, а інколи і логічних каналах зв'язку [9, 10]. Можливість їх взаємного впливу може приводити до істотної зміни параметрів роботи застосунків і є причиною виникнення складних біфуркаційних явищ, які виявляються в нерівномірному або вибуховому характері протікання мережевих процесів (рис. 1).

Тому для управління такими процесами потрібна розробка адекватних математичних моделей, що враховують в конструктивній формі як статистичний характер, так і динаміку передачі пакетів на різних рівнях протоколів міжмережної взаємодії, включаючи часові і просторові характеристики.

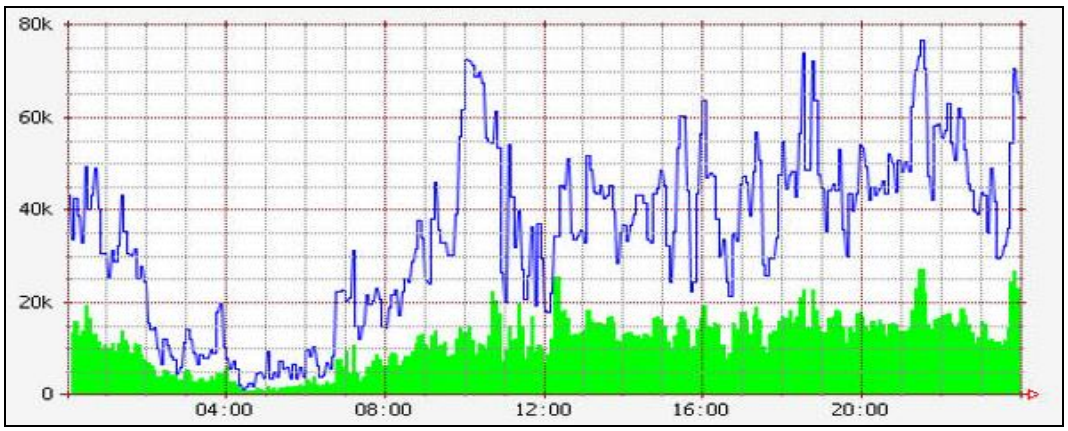


Рис. 1. Вибуховий характер зміни інтенсивності трафіку у віртуальному каналі в залежності від часу

Розгляд прикладів успішного використання мережних технологій дозволяє на якісному рівні визначити властивості комп'ютерних мереж як об'єктів управління і кількісно охарактеризувати різні аспекти використання даних про їх параметри і структуру, взаємодії застосунків, що безпосередньо впливають на якість.

З точки зору інформації про зв'язаність мережевих вузлів, підходи до формування і управління віртуальними з'єднаннями можна розділити на два класи: комутація і маршрутизація. У першому випадку в системі управління аналізуються дані про структуру з'єднання вузлів по всьому маршруту передачі повідомлення, в другому – використовується лише локальна інформація про досяжність суміжних вузлів, а передача даних починається відразу ж після формування запиту. Сучасні підходи до формулювання вимог до мережевих систем управління засновані на складанні специфікацій параметрів віртуальних з'єднань і їх властивостей з врахуванням особливостей застосунків.

Категорія «постійна швидкість передачі» (CBR, Constant Bit Rate) використовується для застосунків, що вимагають жорстких обмежень на величину затримки або її варіації (джітер). Вважається, що для трафіку CBR заздалегідь відомі характеристики потоків даних на канальному і мережному рівнях. Тому можливості CBR перш за все використовують для ізохронних застосунків із впорядкованою доставкою пакетів, наприклад, при передачі аудіо або відео потоків даних. При організації каналу CBR необхідно забезпечити постійну пропускну здатність віртуального з'єднання μ, отже, гарантувати малі зміни у варіаціях затримки ΔT при передачі пакетів. Це досягається шляхом використання методів програмного управління параметрами з'єднанням, в припущенні, що інтенсивність генерації даних λ за час сеансу зв'язку свідомо задовольнятиме умові

$$\lambda(t) < \mu(t). \qquad (1)$$

На практиці, при встановленні з'єднання CBR, через те, що миттєві значення інтенсивності λ, для будь-якого моменту часу вельми важко передбачити заздалегідь, замість умови (2.1) використовується жорсткіше обмеження $\lambda(t) \ll \mu(t)$. В результаті за віртуальним з'єднанням на весь час сеансу зв'язку повністю закріплюються надлишкові ресурси, які заздалегідь узгоджуються з системою управління доступом (Connection admition control, CAC). Проте система CAC може відкинути або відкласти на невизначений час встановлення даного з'єднання CBR, якщо в мережі ATM не вистачає для цього ресурсів.

Категорія «змінна швидкість передачі» (VBR, Variable Bit Rate) застосовується для застосунків, які мають строгі обмеження на величину затримки, але передають дані із змінною інтенсивністю. Прикладом такого типу з'єднань можуть служити компресований голосовий сигнал і цифрові потоки даних, що знімаються з передавальної відео апаратури. Категорія VBR вимагає від віртуального з'єднання забезпечення деякого значення еквівалентної пропускної здатності μ(t), яка була б менше можливого пікового значення, але більше усередненої на всій реалізації інтенсивності генерації даних джерелом трафіку VBR

$$\mu(t) = M\{\lambda(t)\} + b * D, \qquad (2)$$

де $M\{\cdot\}$ – символ операції усереднювання на реалізаціях процесу λ(t); D – параметр, що характеризує дисперсію процесу; b – параметр з'єднання VBR.

У рекомендаціях ATM допустиме значення перевищення пропускної здатності каналу VBR над середнім значенням швидкості генерації даних не регламентується, проте вважається, що її величина повинна вибиратися з врахуванням вірогідності великих локальних відхилень λ(t) від середньостатистичних значень M{λ(t)}. З'єднання, відповідні категорії VBR, організовуються лише в тому випадку, якщо значення еквівалентної пропускної здатності може бути виділене у всіх вузлах мережі, через яку проходить віртуальне з'єднання.

Категорія «невизначена швидкість передачі» (UBR, Unspecify Bit Rate) була розроблена для застосунків, які забезпечують «сервіс найкращих спроб». Категорія UBR характерна для багатьох застосунків, що функціонують в мережі Інтернет і використовують транспортний протокол TCP. Надійність доставки даних і інші характеристики такого віртуального з'єднання не гарантуються жодними механізмами управління на канальному рівні і забезпечуються лише за допомогою транспортних протоколів або засобами самих мережевих застосунків (рис. 2).

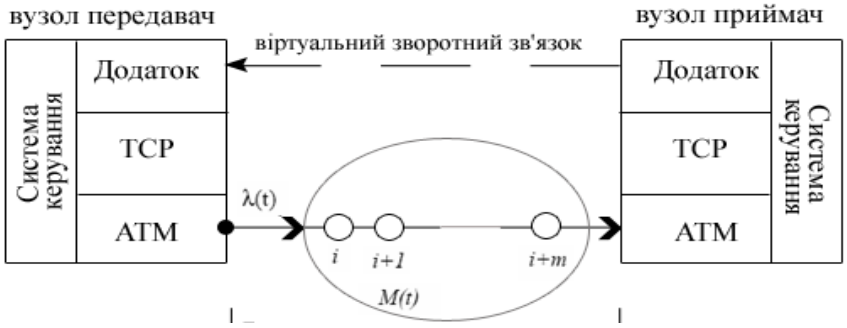


Рис. 2. Структура віртуального з'єднання UBR, що проходить крізь i+*m* вузлів

Вважається, що завжди існує множина потенційних споживачів для даної категорії мережевого сервісу, тому трафік UBR може бути схильний до тимчасових перевантажень, що призводить до втрати даних. Це відбувається, коли число заявок на виділення пропускної здатності з'єднання більше, ніж реально для цих цілей можна надати у момент часу. В результаті підсумкова продуктивність сервісу UBR значною мірою визначається можливостями системи управління прогнозувати стан віртуальних з'єднань і перерозподіляти на цій основі доступні мережеві ресурси.

Досвід роботи з комп'ютерними мережами показав, що методи статистичного мультиплексування дозволяють істотно підвищити рівень використання мережевих ресурсів при збереженні високої якості роботи застосунків за рахунок управління процесами буферизації пакетного трафіку. Найбільшого ефекту можна досягти при зменшенні затримок в ланцюзі зворотного зв'язку за допомогою введення контуру управління інтенсивністю передачі.

Управління трафіком в середовищі мультисервісної мережі

При розгляді механізмів взаємодії різних контурів управління трафіком у високошвидкісних комп'ютерних мережах можна виділити наступні аспекти:

- алгоритми регулювання рівня завантаження в буферах перегружених вузлів;
- методи резервування і перерозподілу ресурсів,
- алгоритми управління швидкістю передачі пакетів в каналах зв'язку.

Всі перераховані вище засоби управління мережевим трафіком мають на меті підвищити продуктивність і якість роботи застосунків, що використовують можливості як транспортного протоколу (наприклад, TCP), так і апаратно-програмних пристроїв мережі ATM для усунення втрат пакетів у віртуальних з'єднаннях. Досвід практичної експлуатації застосунків, що використовують протокол TCP в мережі ATM показує, що без використання спеціальних механізмів оперативного управління перевантаженнями за допомогою відкидання свідомо непотрібних маршрутів, то мережева продуктивність може бути вельми низькою. Тому в алгоритм управління необхідно включити засоби відкидання сегментів TCP (які зазвичай інкапсульовані в IP-пакети), якщо одну із складових пакету, отриманих після процедури фрагментації, було втрачено.

Перший і широко використовуваний в цьому випадку алгоритм управління, заснований на відкиданні всіх складових пакету, наступних за втраченою, яка належить даному пакету IP (PPD, partial packet discard). Модифікацією цього механізму управління є схема раннього відкидання складових (EPD, early packet discard). В цьому випадку використовується алгоритм короткострокового передбачення рівня заповнення буфера комутатора і, якщо довжина черги перевищує певний рівень або деякий критичний поріг, то система управління приймає рішення про відкидання всіх складових, що належать певній кількості (наприклад, одному) сегментів TCP. Така схема управління трафіком дозволяє підвищити продуктивність транспортних з'єднань. Проте, вибір обґрунтованого значення величини критичного порогу заповнення буфера вимагає вивчення такої структури в загальному випадку нестаціонарних потоків даних і створення ефективних алгоритмів передбачення стану каналів.

Друга схема відкидання складових заснована на програмному управлінні ресурсами мережі і вимагає апріорного їх резервування для кожного з віртуальних з'єднань.

Основна перевага такого підходу полягає в тому, що можна гарантувати нульовий рівень втрат складових, викликаних переповнюванням пристроїв буферизації. Проте об'єм ресурсів буферної пам'яті, який дозволяє ефективно використовувати таку схему управління потоком, залежить від величини затримки при поширенні пакетів і флуктуації пропускної здатності лінії зв'язку.

Складність в коректній оцінці необхідного буферного простору призводить до того, що в сучасних комутаторах ATM реалізовано два спрощені варіанти схеми такого управління:

- використання строгого розділення буфера на сегменти, які закріплюються за окремими віртуальними з'єднаннями;
- використання єдиного буферного простору, що розділяється всіма віртуальними з'єднаннями одночасно, об'єм якого може змінюватися в часі.

Хоча теоретично в цьому випадку можливо забезпечити незалежне управління потоками для кожного зі з'єднань, але при цьому вельми неефективно використовується найбільш дорогий ресурс комутатора – пам'ять, яка закріплена за окремим буфером.

Для подолання подібного недоліку застосовується модифікована схема спільного використання вільного об'єму буфера всіма віртуальними каналами. При такій схемі управління істотні переваги мають алгоритми із зворотним зв'язком, безпосередньо реалізованим в кожному з каналів зв'язку.

Для цього за віртуальним каналом закріплюється певна кількість доступних для використання узагальнених мережевих ресурсів (кредитів). Пакет з конкретного віртуального з'єднання може бути переданий лише в тому випадку, якщо це з'єднання має позитивну кількість кредитів. Інформація про величини кредитів пересилаються по мережі одночасно з передачею пакетів даних між вузлами. У новому вузлі число кредитів змінюється відповідно до його поточного стану.

Практична реалізація такої кредитної схеми передбачає, що буфер в передавальному вузлі розділяється на дві зони.

Висновки

В даній статті розглянуто класифікації мережевих характеристик та були визначені спеціальні категорії трафіку, для яких використовуються різні системи управління. Проведено аналіз методів управління, виявлено основні переваги, а також проблеми даних методів, що вимагають подальшого їх дослідження.

Напрямком подальших досліджень є розробка адекватних моделей трафіку та відповідне вдосконалення методів управління процесами передачі даних.

СПИСОК ЛІТЕРАТУРИ

1. Ярошевич. Р.О. Аналіз підходів до мінімізації затримок тактильного інтернету у комп’ютерних мережах / Р.О. Ярошевич, А.А. Коваленко // Проблеми інформатизації. Тези доповідей дев’ятої міжнародної НТК. – Черкаси: ЧДТУ; Харків: НТУ «ХПІ»; Баку: ВА ЗС АР; Бельсько-Бяла: УТіГН; Харків: ДП «ПД ПКНДІ АП», 2021. – 18-19 листопада 2021. – С. 101.
2. Tehrani M. N., et al. Device-to-Device Communication in 5G Cellular Networks: Challenges, Solutions, and Future Directions // IEEE Communications Magazine. 2014. Vol. 52. Iss. 5. pp. 86–92.
3. Ateya A. A., Vybornova A., Kirichek R., Koucheryavy A. Multilevel Cloud Based Tactile Internet System // Proc. of 19th International Conference on Advanced Communication Technology (ICACT) 2017. pp. 105–110
4. Volkov, A., Ateya, A. A., Muthanna, A., Kirichek, R. MEC and SDN/NFV as a Solution for Providing 1ms in 5G/IMT-2020 Communication Networks // 73rd All-Russian Scientific-Technical Conference, Dedicated to the Day of Radio. 2018. Pp. 192–193.
5. Kovalenko, A., Kuchuk, H., Kuchuk, N. and Kostolny, J. (2021), “Horizontal scaling method for a hyperconverged network”, *2021 International Conference on Information and Digital Technologies* (IDT), Zilina, Slovakia, doi: https://doi.org/10.1109/IDT52577.2021.9497534
6. Kuchuk, N., Kovalenko, A., Ruban, I., Shyshatskyi, A., Zakovorotnyi, O. And Sheviakov, I. (2023), “Traffic Modeling for the Industrial Internet of NanoThings”, *2023 IEEE 4th KhPI Week on Advanced Technology, KhPI Week 2023 - Conference Proceedings*, 194480, doi: http://dx.doi.org/10.1109/KhPIWeek61412.2023.10312856
7. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об’єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: https://doi.org/10.20998/2522-9052.2018.1.04
8. Зиков І. С., Кучук Н. Г., Шматков С. І. Синтез архітектури комп'ютерної системи управління транзакціями e-learning. *Сучасні інформаційні системи*. 2018. Т. 2, № 3. С. 60–66. DOI: https://doi.org/10.20998/2522-9052.2018.3.10
9. Ruban, I.V., Martovytskyi, V.O., Kovalenko, A.A. and Lukova-Chuiko, N.V. (2019), “Identification in Informative Systems on the Basis of Users' Behaviour”, Proceedings of the International Conference on Advanced Optoelectronics and Lasers, CAOL 2019-September,9019446, pp. 574-577, DOI: https://doi.org/10.1109/CAOL46282.2019.9019446
10. Sharma, S.K., Woungang, I., Anpalagan, A., Chatzinotas, S. (2020). Toward Tactile Internet in beyond 5G era: recent advanes current issues, and future directions. Ieee Access, 8, 56948-56991.

Received (Надійшла) 22.11.2023
Accepted for publication (Прийнята до друку) 17.01.2024

Analysis of methods for data transfer processes and traffic management in multiservice computer networks

Viktor Ganzii, Andriy Kovalenko, Oleksii Sytnyk

Abstract. The purpose of this work is to analyze the methods for data transfer processes and traffic management that will be able to increase the efficiency of data transmission in multiservice computer networks. The growing needs of society for new services of telecommunication networks lead to a change in the ideology of the construction of the latter every decade. Today, technologies using wavelength division multiplexing and compression are being replaced by multiservice technologies, the basic concept of which is to separate transport and switching functions, transaction management functions, and service management functions from each other. In particular, it can be diverse corporate networks where it is important to control access to resources, cloud environments where various services are provided, online game services where low latency and stable connection are required. In order to increase the efficiency of multiservice computer networks today, technical means using asynchronous transmission mode (ATM or B-ISDN technology) are integrated with applications for the Internet. When building such networks, it is important to consider the classification of network characteristics, namely the category of traffic for choosing the optimal management method. In this article, the method of software control for connection parameters and methods of statistical multiplexing of data transmission are considered, the main advantages and problems of these methods, which require further research, are analyzed.

Keywords: method, management, process, multiservice computer network, model, interaction, traffic, application.

УДК 621.3 **doi: 10.26906/SUNZ.2024.1.193**

К. М. Лейченко, Г. В. Фесенко

Національний аерокосмічний університет "Харківський авіаційний інститут", Харків, Україна

ПРОГРАМНИЙ ЗАСІБ ПІДТРИМКИ ПЛАНУВАННЯ РОЗГОРТАННЯ LiFi МЕРЕЖІ НА ОСНОВІ БПЛА ДЛЯ ЗАБЕЗПЕЧЕННЯ ПЕРЕДАЧІ ДАНИХ В УМОВАХ РУЙНУВАНЬ

Анотація. Аварії на об'єктах критичної інфраструктури супроводжуються пошкодженням штатних мереж передачі даних від датчиків контролю критично важливих параметрів технологічного обладнання до кризових центрів (КЦ). Відсутність таких даних може призводити до помилкових і недостатньо обґрунтованих рішень з боку персоналу КЦ під час дій по локалізації та ліквідації наслідків аварії. У якості альтернативи пошкодженим штатним мережам може розглядатися LiFi мережа на основі безпілотних літальних апаратів (БПЛА), де останні виступають у ролі ретрансляторів. Однак, внаслідок руйнувань обладнання і конструкцій у виробничих приміщеннях можуть утворюватися механічні перешкоди, які потребуватимуть прокладання маршрутів розповсюдження LiFi сигналу в обхід цих перешкод. Предметом статті є засоби планування розгортання літаючих мереж для забезпечення передачі даних в умовах руйнувань. Мета статті – запропонувати програмний засіб підтримки планування розгортання LiFi мережі на основі БПЛА для забезпечення передачі даних в умовах руйнувань приміщень об'єктів критичної інфраструктури. Завдання статті: запропонувати схему розгортання LiFi мережі на основі БПЛА у виробничому приміщенні з перешкодами; представити архітектуру і особливості застосування програмного засобу для реалізації запропонованої схеми; надати приклади використання програмного засобу. Отримані наступні результати роботи. Розроблено програмний засіб для підтримки планування розгортання LiFi мережі на основі БПЛА у приміщеннях з перешкодами, який дозволяє прокладати маршрути розповсюдження LiFi сигналу (світлового потоку з даними) і визначати для його передачі необхідну кількість БПЛА і місця їхнього розміщення на прокладеному маршруті. Надані результати застосування програмного засобу для планування розгортання LiFi мережі в заданому приміщенні з перешкодами з використанням для їх обходу і побудови маршрутів методів прямокутника і керованого водоспаду. Напрямок подальших досліджень полягає у розробці методу і програмного засобу визначення необхідної кількості змін і чисельності БПЛА у кожній з них для забезпечення безперебійної роботи утвореної ними LiFi мережі протягом заданого часу з визначеною замовником ймовірністю безвідмовної роботи.

Ключові слова: безпілотний літальний апарат, LiFi мережа, програмний засіб, обхід перешкод, прокладання маршруту.

Вступ

Постановка проблеми. Аварії на об'єктах критичної інфраструктури супроводжуються пошкодженням штатних мереж передачі даних від датчиків контролю критично важливих параметрів технологічного обладнання до кризових центрів (КЦ). Відсутність таких даних може призводити до помилкових і недостатньо обґрунтованих рішень з боку операторів КЦ під час їх дій по локалізації та ліквідації наслідків аварії. У якості альтернативи пошкодженим штатним мережам може розглядатися LiFi мережа [1] на основі безпілотних літальних апаратів (БПЛА), де останні виступають у ролі ретрансляторів. Технологія LiFi може забезпечувати дуже високі швидкості передачі даних, що перевершують багато традиційних технологій, таких як Wi-Fi. Це особливо корисно у ситуаціях, де потрібна висока пропускна спроможність. Крім того ця технологія не створює електромагнітних перешкод, які можуть впливати на інші бездротові мережі, та забезпечує високий рівень безпеки передачі даних, оскільки світловий потік не виходить за межі виробничого приміщення. Що ж стосується БПЛА, то їх використання для утворення LiFi мережі обумовлено тим, що вони забезпечують гнучкість у виборі розташування для передачі даних, що може бути критично в умовах руйнувань.

Однак, внаслідок руйнувань обладнання і конструкцій у виробничих приміщеннях можуть утворюватися механічні перешкоди, які потребуватимуть прокладання маршрутів розповсюдження LiFi сигналу в обхід цих перешкод. Моделювання різних сценаріїв прокладання маршрутів розповсюдження LiFi сигналу у виробничих приміщеннях з перешкодами дозволить більш обґрунтовано здійснювати планування розгортання LiFi мережі на основі БПЛА для забезпечення передачі даних в умовах руйнувань.

Аналіз останніх досліджень і публікацій. Питання моделювання роботи БПЛА розглядаються у багатьох роботах. Наприклад, у [2] наголошується, що моделювання є важливим інструментом для розробки та тестування БПЛА. Моделювання дає змогу інженерам досліджувати поведінку БПЛА в різних умовах, що може допомогти їм уникнути дорогих і небезпечних випробувань на реальних БПЛА.

Автори [3] представили комп'ютерний симулятор для моделювання поведінки БПЛА в середовищі з перешкодами. При цьому вони використовували планування на основі графів, управління на основі зворотного зв'язку та різні методи машинного навчання. У роботі [4] розглянуто реалізацію наземної станції управління (НСУ) для симулятора польоту БПЛА. НСУ дає змогу оператору керувати БПЛА в симуляторі та спостерігати за його поведінкою. У [5] за допомогою обчислювального моделювання досліджується вплив запуску боєприпасів на стійкість шестироторних БПЛА. Автори аналізують вплив різних чинників, таких як кут запуску боєприпасів, час запуску, положення установки і маса, на стійкість БПЛА. У статті [6] розглядається розробка наземного симулятора польоту з регульованою стійкістю для пілотної підготовки. Такий симулятор дає змогу імітувати

поведінку літака в різних умовах польоту, включно зі змінами аеродинамічних характеристик і відмовами систем керування. Це допомагає пілотам тренувати свої навички пілотування і справлятися з несподіваними ситуаціями. Автори [7] обговорюють ключові компоненти та функції програмного забезпечення, необхідні для безпечного та ефективного керування БПЛА. Робота [8] пропонує структуру для планування ефективних експериментів з роями БПЛА на основі симулятору. Автори підкреслюють важливість вибору симулятору за такими параметрами, як масштабованість, достовірність та зручність використання. У публікації [9] розглядається фреймворк UTSim, який призначений для інтеграції БПЛА до керування повітряним рухом.

Однак у проаналізованих роботах не розглядаються випадки реалізації програмного засобу моделювання сценаріїв застосування БПЛА для утворення LiFi мереж всередині приміщення.

Метою роботи є запропонувати програмний засіб підтримки планування розгортання LiFi мережі на основі БПЛА для забезпечення передачі даних в умовах руйнувань приміщень об'єктів критичної інфраструктури. Програмний засіб використовує попередньо розроблені авторами цієї статті методи прямих кутів [10] та керованого водоспаду [11].

Розробка схеми планування розгортання LiFi мережі на основі БПЛА у виробничому приміщенні з перешкодами

Схему планування розгортання LiFi мережі на основі БПЛА у виробничому приміщенні з перешкодами подано на рис. 1. Планування відбувається у три етапи.

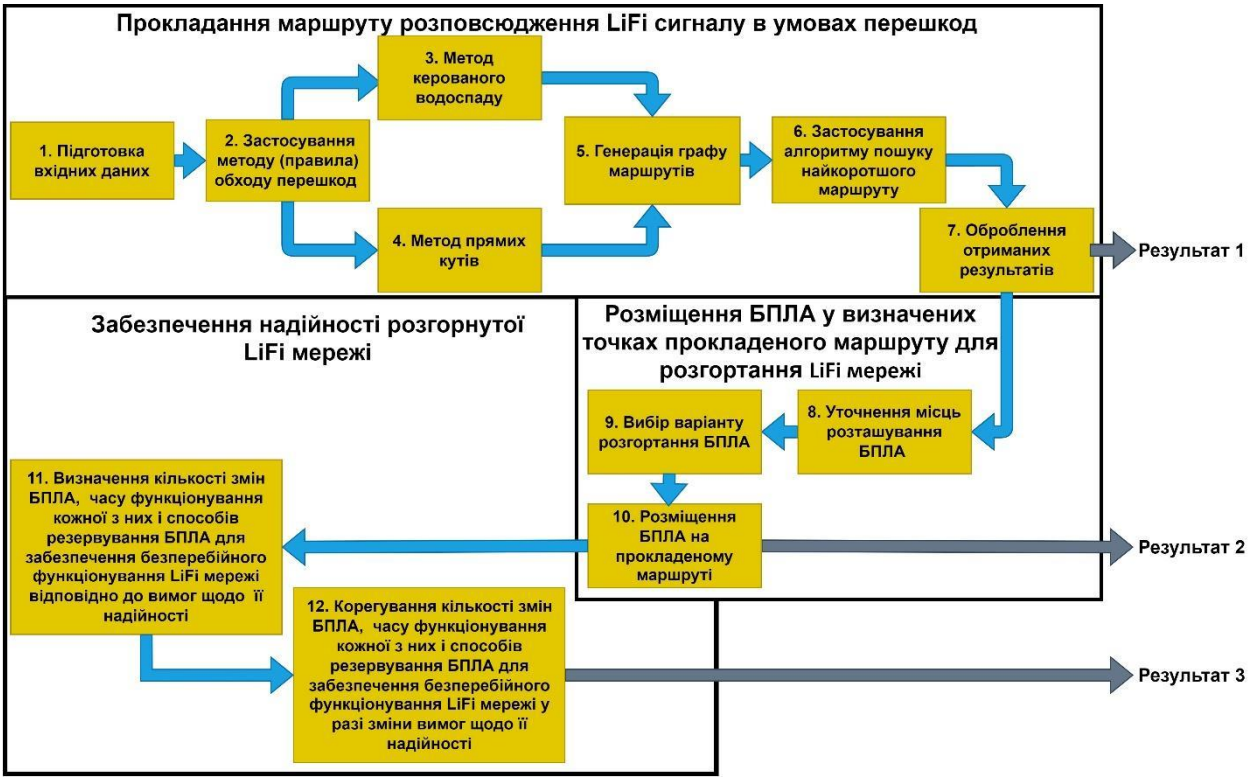


Рис. 1. Схема планування розгортання LiFi мережі на основі БПЛА у виробничому приміщенні з перешкодами

Етап 1. *Прокладання маршруту розповсюдження LiFi сигналу в умовах перешкод*. У найпростішому варіанті завдання може бути сформульовано так: потрібно прокласти маршрут (лінію LiFi зв'язку) з точки *A* (джерело інформації) в точку *B* (користувач інформації) у виробничому приміщенні з перешкодами в двовимірному (2D) просторі. На початку необхідно оцінити наслідки руйнувань у виробничому приміщенні і скласти карту перешкод, причому для реалізації розглянутих далі методів кожну з перешкод представляємо у вигляді прямокутника. Далі обираємо один з методів обходу перешкод: метод прямих кутів (коли обхід кожної перешкоди здійснюється виключно за правилом лівого або правого кута) або метод керованого водоспаду. При реалізації кожного з методів приймаємо припущення про те що точки *A* і *B*, а також перешкоди є статичними і не змінюються з часом. Застосування зазначених методів дозволяє сформувати множину маршрутів від точки *A* до точки *B* у 2D просторі виробничого приміщення, а поєднання всіх точок маршруту (*A*, *B* і точок зміни напрямку руху за маршрутом внаслідок обходу перешкод) – згенерувати граф маршрутів. Наявність такого графа на наступному кроці дозволяє застосувати алгоритм пошуку найкоротшого маршруту (наприклад, алгоритм Дейкстри) розповсюдження LiFi сигналу в умовах перешкод від джерела інформації (точка *A*) до її споживача (точка *B*). Заключним кроком першого етапу є оброблення отриманих результатів з метою їх подальшого використання на етапах 2 і 3.

Етап 2. *Розміщення БПЛА у визначених точках прокладеного маршруту для розгортання LiFi мережі.* Першим кроком на цьому етапі є визначення переліку варіантів руху кожного БПЛА з місця свого

базування до заданої точки маршруту для розгортання LiFi мережі. На наступному кроці відповідно із заданим критерієм (наприклад, часом розгортання) обирається кращій варіант і здійснюється розміщення БПЛА у визначених точках прокладеного маршруту.

Етап 3. *Забезпечення надійності розгорнутої LiFi мережі.* У разі отримання вимог від замовника стосовно мінімально необхідного значення імовірності безвідмовної роботи LiFi мережі, визначається спосіб її резервування та кількість резервних БПЛА. У разі зміни вимог, здійснюється коригування способу резервування та/або кількості резервних БПЛА.

Програмний засіб

Архітектура програмного засобу. Програмний засіб, що пропонується і має назву "Simulation Way", може бути використаний для реалізації етапів 1 та 2 схеми планування розгортання LiFi мережі на основі БПЛА, представленої на рис. 1. Він має трьохрівневу архітектуру (рис. 2).

1) Рівень *GUI*. Цей рівень являє собою інтерфейс програми, який реалізований за допомогою Python бібліотеки tkinter.

2) Рівень *Business logic*. Цей рівень надає логіку взаємодії між сховищем даних (*Storage data*) та графічним інтерфейсом (*GUI*). Рівень містить модулі генерування звітів (результатів), модулі розрахункового ядра та модулі зовнішніх алгоритмів (наприклад алгоритм Дейкстри для пошуку найкоротшого маршруту), які можуть взаємодіяти між собою.

3) Рівень *Storage data*. Цей рівень відповідає за отримання і зберігання у вигляді файлів даних результатів розрахунків та звітів щодо процесу роботи програмного засобу.

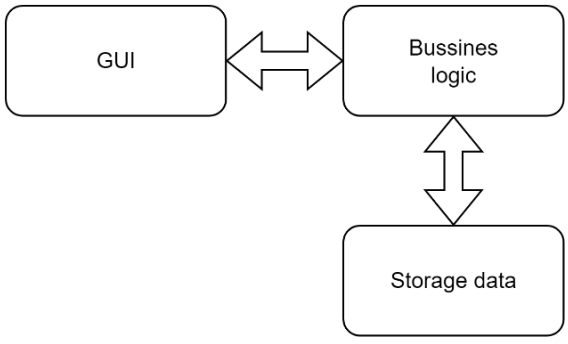


Рис. 2. Архітектура програмного засобу "Simulation Way"

Варіанти використання програмного засобу. Програмний засіб може бути використаний операторами КЦ під час моделювання розгортання LiFi мережі на основі БПЛА у виробничому приміщенні з перешкодами. Варіанти використання програмного засобу оператором КЦ продемонстровано на рис. 3.

Взаємодія оператора КЦ з програмним засобом здійснюється за допомогою графічного інтерфейсу.

Оператору КЦ доступна велика кількість налаштувань: встановлювати розміри робочої площі виробничого приміщення, генерувати необхідну кількість перешкод з заданими характеристиками та обирати методи (правила) обходу цих перешкод. Крім того, сформовані вхідні дані під час однієї ітерації моделювання можуть використовуватися у наступних ітераціях.

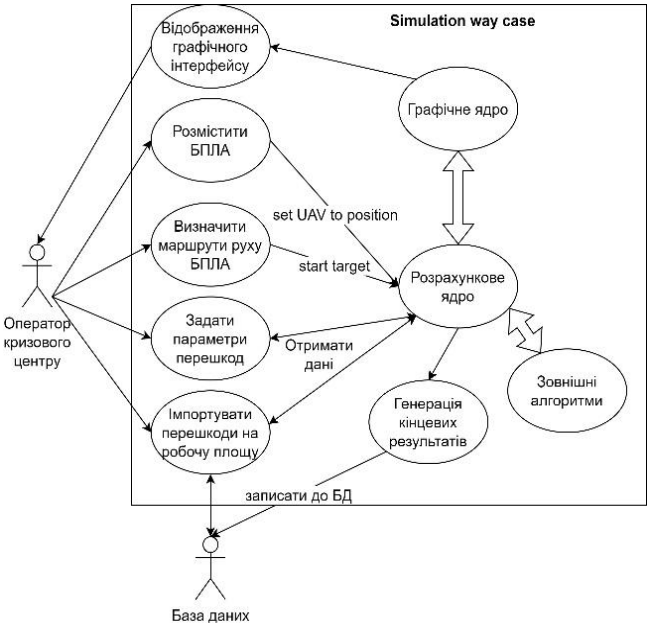


Рис. 3. Діаграма варіантів використання програмного засобу "Simulation Way"

За проведення розрахунків відповідає *Розрахункове ядро*, яке поєднує у собі математичні та алгоритмічні модулі, а також модулі взаємодії. *Графічне ядро* інформує оператора КЦ про хід процесу моделювання, виводить результати розрахунків та дозволяє керувати правилами моделювання. Зокрема, оператор КЦ візуально може бачити:

- номер ітерації;
- метод (правило) обходу перешкод;
- координати обраних ним точок на прокладеному маршруті;
- довжину маршруту (довжину заданого відрізку маршруту);
- кількісті БПЛА, які потрібно розташувати у визначених точках маршруту для розгортання LiFi мережі.

Послідовність взаємодії оператора КЦ з програмним засобом показана на рис. 4.

Під час активації програмного засобу ініціалізуються необхідні бібліотеки та модулі. Оператор КЦ напряму не взаємодіє з алгоритмами або *Розрахунковим ядром*. Для цього оператор КЦ має у своєму розпорядженні графічний інтерфейс. Функції оператора КЦ під час користування програмним засобом зводяться до введення (корегування) необхідних для моделювання параметрів. На підставі згенерованих під час моделювання множини маршрутів, програмний засіб формує граф маршрутів і дозволяє активувати зовнішній алгоритм, наприклад алгоритм Дейкстри, для визначення найкоротшого маршруту від джерела інформації до її споживача (від початкової до кінцевої точки маршруту). Отримання результатів розрахунків можливе як у вигляді візуальної інформації на панелі керування графічного інтерфейсу, так і у вигляді підготовленого файлу звіту.

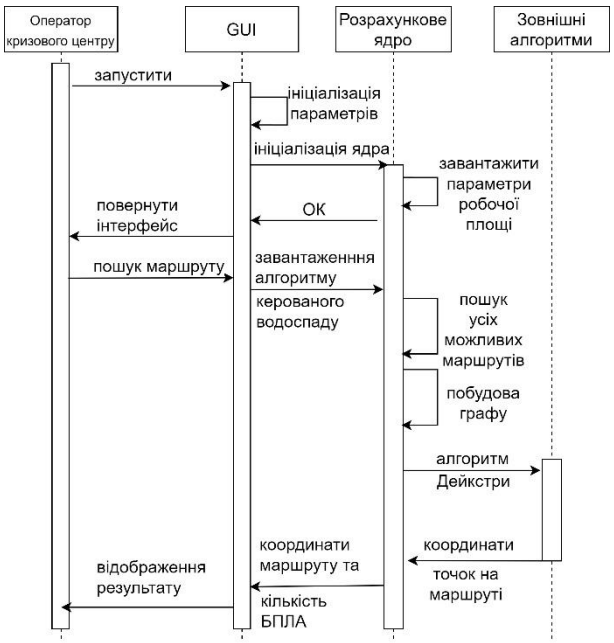


Рис. 4. Діаграма, що демонструє послідовність взаємодії оператора КЦ з програмним засобом “Simulation Way”

Структура програмного засобу. Програмний засіб “Simulation Way” має модульну структуру, де кожен модуль реалізує API для взаємодії, відокремлюючи окремі функціональні блоки. На рис. 5 зображена модульна діаграма, яка демонструє логіку взаємодії компонентів програмного засобу між собою.

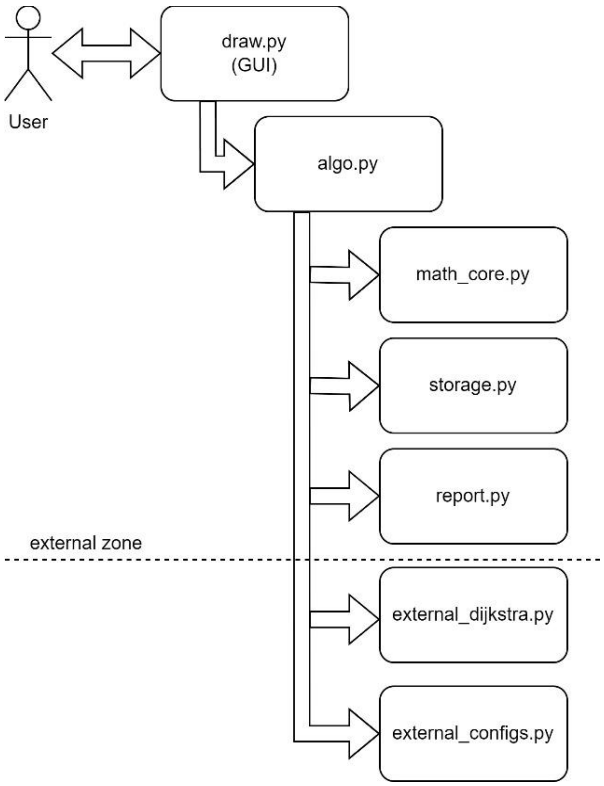


Рис. 5. Модульна діаграма, яка демонструє логіку взаємодії компонентів програмного засобу “Simulation Way” між собою

Оператор КЦ взаємодіє з програмним засобом за допомогою модуля *draw.py*. Графічний інтерфейс (GUI) реалізовано за допомогою Python бібліотеки *tkinter*, яка є одним із найпоширеніших рішень для відображення графічного інтерфейсу. Блок *algo.py* реалізує міжмодульну взаємодію (обхід перешкод, рух БПЛА тощо) та надає базовий функціонал. Модуль *math_core* є математичним ядром програмного засобу. У цьому модулі реалізоване API для розрахунку довжин, конвертації та корекції чисел з плаваючою крапкою тощо. *External zone* представляє собою зону зовнішніх модулів, які використовуються для взаємодії з ресурсами операційної системи та базовими інтегрованими алгоритмами. Саме завдяки *External zone* стає можливим:

- генерувати зображення графів при використанні різних методів (правил) обходу перешкод;
- генерувати дані про: параметри перешкод, кількість ітерацій, довжину маршруту (ділянки маршруту), кількість БПЛА, необхідних для розгортання LiFi мережі, координати точки розташування кожного БПЛА на прокладеному маршруті;
- вести журнал роботи програмного засобу.

Модуль *storage.py* представляє собою сховище для змінних, які необхідні для міжмодульної взаємодії. Цей модуль зберігає характеристики БПЛА, координати їх місць базування та точок подальшого розташування на маршруті, координати розташування перешкод тощо. Дані, які формуються у процесі використання програмного засобу першочергово зберігаються саме у цьому модулі (координати точок маршруту для конкретної ітерації, методи (правила) обходу перешкод тощо)

Опис базового функціоналу. Графічний інтерфейс програмного засобу “Simulation Way” розташований в окремих графічних вікнах: *Control* (рис. 6, 7), яке є центром керування параметрами та правилами моделювання (рис. 7), та *Way simulation* (рис. 8), яке відображає хід процесу моделювання у реальному часі.

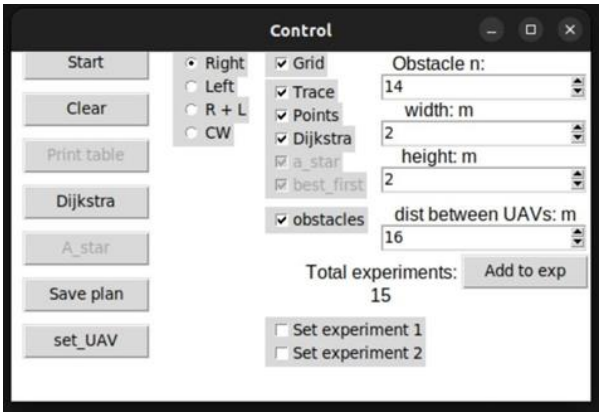


Рис. 6. Вид панелі *Control*

Вікно інтерфейсу, представлене на рис. 7, умовно можна поділити на чотири зони.

1) *Зона запуску процесів*. Кнопка *Start* відповідає за запуск процесу пошуку маршруту між початковою та кінцевою точками за заданими параметрами. Кнопка *Clear* дозволяє очистити дані з попередніх ітерацій. Кнопка *Dijkstra* дозволяє використати алгоритм Дейкстри для пошуку найкоротшого маршруту за поточним графом. Кнопка *Start* у поєднанні з пунктом

CW із зони 2 також використовує цей алгоритм, проте граф формується автоматично та не може бути змінений протягом усіх ітерацій моделювання. Кнопка *Save plan* відповідає за збереження поточного плану приміщення разом з перешкодами. Дані будуть збережені у xml-файл та можуть бути повторно використані у майбутньому. Кнопка *set_UAV* відповідає за розміщення БПЛА в заданих точках прокладеного маршруту.

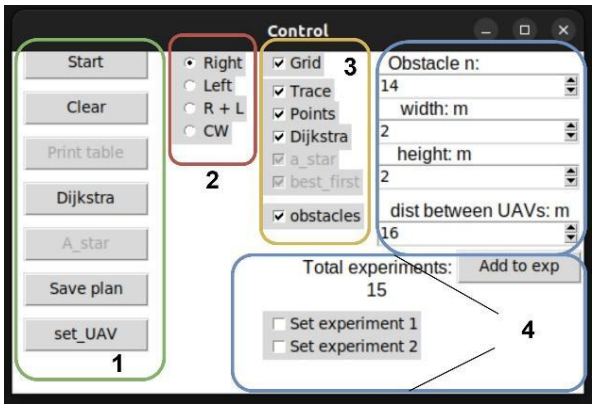


Рис 7. Вид панелі Control з функціональними зонами: 1 – зона запуску процесів; 2 – зона реалізації методу (правила) обходу перешкод; 3 – зона графічного відображення шарів; 4 – зона налаштування параметрів моделювання

2) *Зона реалізації методу (правила) обходу перешкод*. Ця зона відповідає за реалізацію методу (правила) обходу перешкоди у виробничому приміщенні. У разі вибору *Right* обхід перешкоди буде за правилом правого кута, а *Left* – лівого кута. Вибір *R+L* дозволить використовувати два попередніх правила, але кожен раз обхід перешкоди буде здійснюватися через кут, до вершини якого відстань під час обходу є найменшою.

3) *Зона графічного відображення шарів*. Ця зона дозволяє включити та виключити кожен шар без втрати даних та обмежень.

4) *Зона налаштування параметрів моделювання*. Ця зона дозволяє встановити кількість перешкод, їх форму та автоматизувати процес створення статистичних даних у звіті, який формується для оператора КЦ.

Панель *Way simulation* (рис. 8) відображає робочу площу виробничого приміщення у 2D просторі.

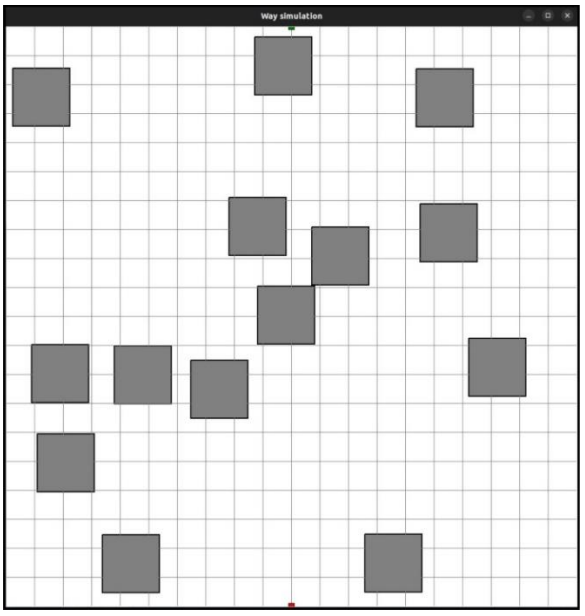


Рис 8. Приклад відображення робочої площі виробничого приміщення у 2D просторі на початку моделювання

Маленькі зелений і червоний прямокутники на рис. 8 є відповідно початковою (джерело інформації) та кінцевою (споживач інформації) точками маршруту. На робочій площі знаходяться згенеровані перешкоди, які зображені для демонстрації у вигляді прямокутника (за наявності більш складних форм перешкод їх проекція може бути вписана у випуклий багатокутник з довільною кількістю кутів).

Якщо обрати для обходу перешкод метод керованого водоспаду, то програмний засіб згенерує граф шляхів, представлений на рис. 9. Далі цей граф може бути використаний для реалізації зовнішнього алгоритму – алгоритму Дейкстри (*Dijkstra*).

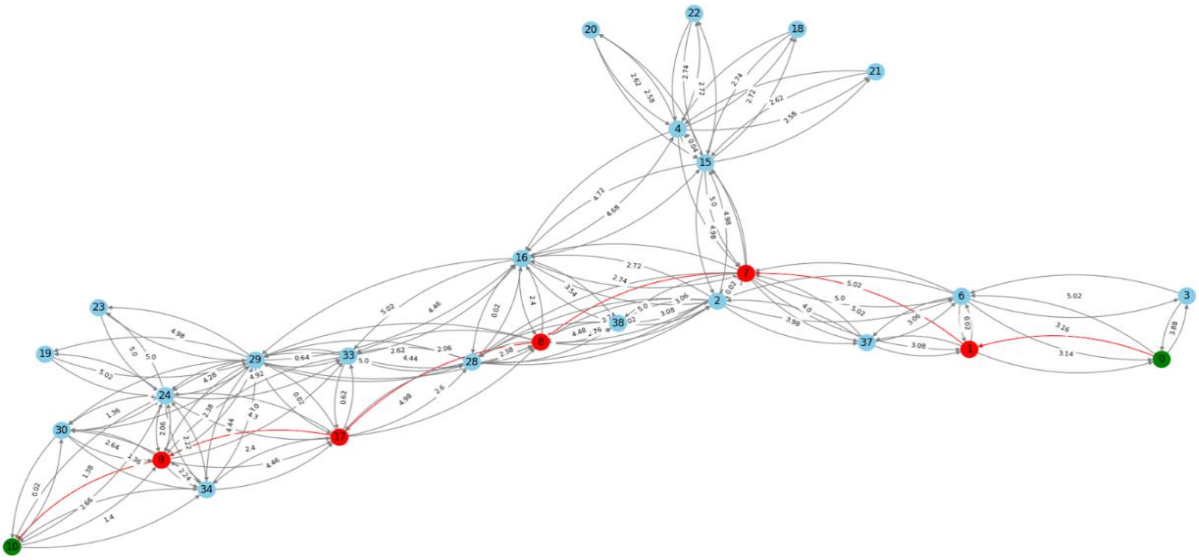


Рис. 9. Згенерований граф маршрутів для методу керованого водоспаду

Зеленим кольором на графі показані вершини, що відповідають початковій та кінцевій точкам найкоротшого прокладеного маршруту, а червоним – його проміжні точки.

Решта вершин показано синім кольором. Кожному ребру у відповідність поставлено його вагу, яка означає відстань між вершинами (точками маршруту) у метрах.

Приклади використання

Послідовно розглянемо приклади використання програмного засобу “Simulation Way” для прокладання маршруту розповсюдження LiFi сигналу з використанням для обходу перешкод правил лівого та правого кута, а також методу керованого водоспаду.

Для використання правила правого кута для обходу перешкод необхідно на панелі *Control* виставити перемикач на позицію *Right* та натиснути *Start* (рис. 10).

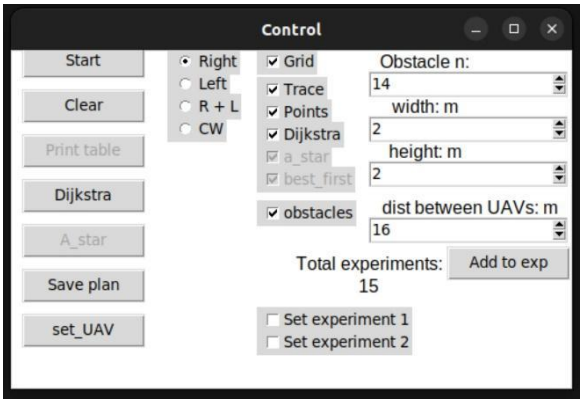


Рис. 10. Панель *Control* з параметрами для прокладання маршруту розповсюдження LiFi сигналу з використанням для обходу перешкод правила правого кута

Результат моделювання представлено на рис. 11. Зелена ламана лінія, показна на цьому рисунку, і є прокладений маршрут. Маленькими зеленими прямокутниками показані початкова і кінцева точки.

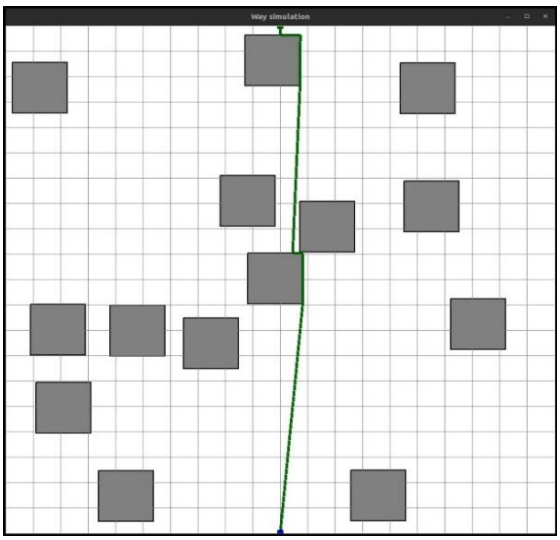


Рис. 11. Панель із зображенням прокладеного маршруту розповсюдження LiFi сигналу з використанням для обходу перешкод правила правого кута

Крім того, програмний засіб визначає та зберігає координати точок, які у майбутньому будуть використані як точки розміщення БПЛА у складі утвореної ними LiFi мережі. У якості таких точок розглядають усі вершини ламаної (маршруту) на рис. 11, окрім початкової (джерело інформації) і кінцевої (споживач інформації).

Для активації процесу моделювання з використанням для обходу перешкод правила лівого кута необхідно на панелі *Control* виставити перемикач на позицію *Left* та натиснути *Start* (рис. 12).

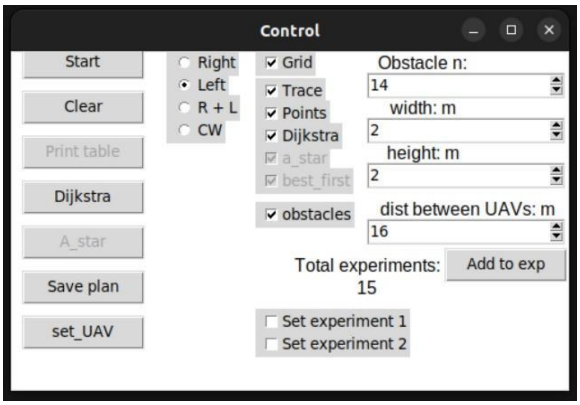


Рис. 12. Панель Control з параметрами для прокладання маршруту розповсюдження LiFi сигналу з використанням для обходу перешкод правила лівого кута

Як і у попередньому випадку результатом моделювання буде згенерований маршрут розповсюдження LiFi сигналу у вигляді зеленої ламаної лінії (рис. 13).

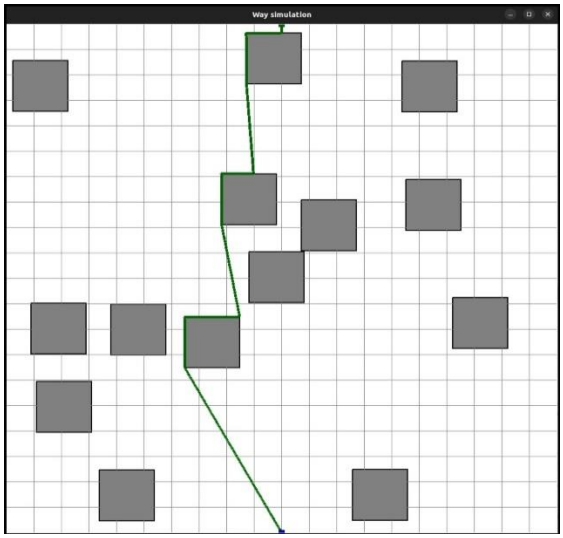


Рис. 13. Панель із зображенням прокладеного маршруту розповсюдження LiFi сигналу з використанням для обходу перешкод правила лівого кута

Для активації процесу моделювання з використанням для обходу перешкод методу керованого водоспаду, необхідно на панелі *Control* виставити перемикач на позицію *CW* та натиснути *Start* (рис. 14). Метод керованого водоспаду дещо відрізняється від раніше розглянутих правил лівого та правого кутів, які є похідними від метода прямих кутів. Цей метод

складається з декількох кроків, кожен з яких має своє завдання.

Перший крок має на меті сформувати граф усіх можливих маршрутів. Для цього будується шлях за допомогою методу прямих кутів, де одночасно застосовуються правила лівого і правого кутів. Це дозволить побудувати граф всіх можливих маршрутів і використати на ньому алгоритм пошуку найкоротшого маршруту, наприклад, алгоритм Дейкстри.

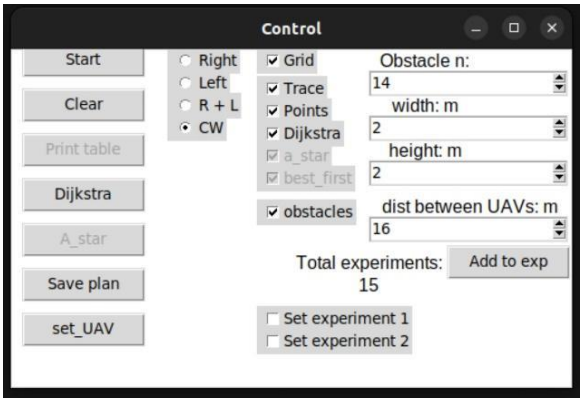


Рис. 14. Панель Control з параметрами для прокладання маршруту розповсюдження LiFi сигналу з використанням для обходу перешкод методу керованого водоспаду

Маршрут буде відображатися червоною ламаною лінією з зеленими точками, які позначають місця для розміщення БПЛА на прокладеному маршруті (рис. 15).

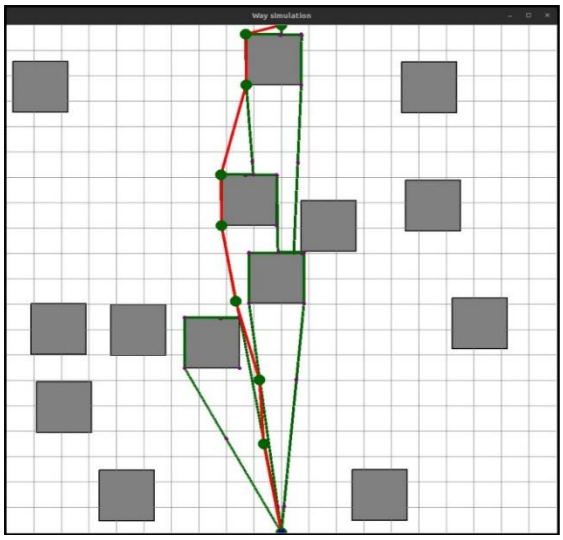


Рис. 15. Панель із зображенням прокладеного маршруту розповсюдження LiFi сигналу з використанням для обходу перешкод методу керованого водоспаду

Як ми можемо бачити, тут точками майбутнього розміщення БПЛА для утворення LiFi мережі будуть не тільки вершини ламаної лінії, а й певна кількість додаткових точок. Необхідність їх введення обумовлена тим, що в умовах запиленості і задимленості виробничого приміщення внаслідок аварійних руйнувань і загорянь відстань між вершинами ламаної лінії може не забезпечувати задану якість передачі даних світловим потоком. Таким чином, між БПЛА, які розташуються на точках сусідніх вершин, необхідно буде розміщувати додатковий (додаткові) БПЛА.

Для генерації маршрутів руху БПЛА з місць базування до точок розміщення на прокладеному маршруті розповсюдження LiFi сигналу необхідно на панелі *Control* натиснути кнопку *set_UAV*. Для прокладання маршрутів руху БПЛА можуть використовуватися ті ж самі методи (правила), що для прокладання маршруту розповсюдження LiFi.

У прикладі, представленому на рис. 16, для обходу перешкод використовується одночасно правила лівого та правого кутів (перемикач поставлено у положення *R+L*).

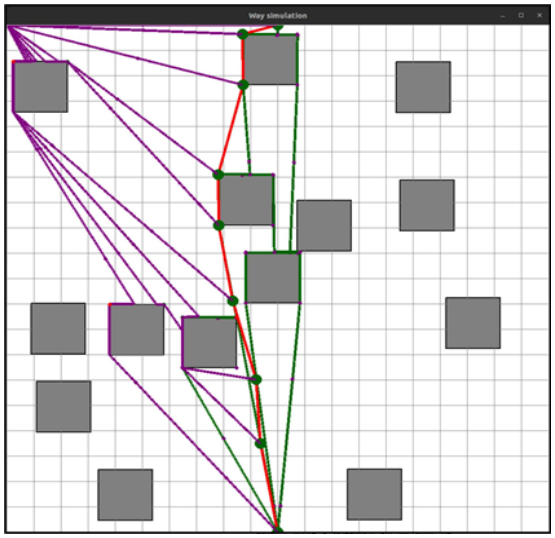


Рис. 16. Панель із зображенням прокладених маршрутів руху БПЛА з місць базування до точок розміщення на прокладеному маршруті розповсюдження LiFi сигналу з одночасним використанням для обходу перешкод правил лівого та правого кутів

На рис. 16 фіолетовими лініями позначено маршрути руху кожного БПЛА до місця розташування на прокладеному маршруті розповсюдження LiFi сигналу.

Приклад на рис. 16 є демонстраційним і точки, до яких рухаються БПЛА, вважаються заздалегідь визначеними. У загальному випадку точка розміщення для кожного БПЛА може визначатися за різними критеріями – ресурс бортової батареї, час розгортання LiFi мережі тощо. У разі розгортання LiFi мережі для роботи протягом тривалого часу можуть опрацьовуватися стратегії позмінного чергування груп БПЛА у складі мережі з визначенням маршрутів руху БПЛА до точок розміщення на маршруті і повернення з нього до місць базування для підзарядки.

Висновки

У статті представлено програмний засіб “Simulation Way”, який може бути використаний операторами КЦ об’єкту критичної інфраструктури для моделювання різних сценаріїв розгортання LiFi мережі на основі БПЛА у виробничому приміщенні з перешкодами. Програмний засіб має трьохрівневу архітектуру, у якій функціональні блоки виконані у вигляді модулів, які можуть взаємодіяти між собою в процесі

моделювання. Програмний засіб має зручний інтерфейс і дозволяє вирішувати такі основні завдання:

- генерувати перешкоди із заданими параметрами;
- прокладати маршрути розповсюдження LiFi сигналу (світлового потоку з даними) у виробничому приміщенні з перешкодами з використанням методу прямих кутів (правила лівого (правого) кута) та методу керованого водоспаду;
- формувати граф можливих маршрутів і застосовувати на ньому алгоритм Дейкстри для пошуку найкоротшого маршруту від джерела інформації до її споживача;
- визначати точки розташування БПЛА на прокладеному маршруті розповсюдження LiFi сигналу для утворення LiFi мережі;
- визначати маршрути руху БПЛА з місць базування до визначених точок на прокладеному маршруті розповсюдження LiFi сигналу.

Подальші дослідження можуть бути спрямовані на розроблення методу і програмного засобу визначення необхідної кількості змін і чисельності БПЛА у кожній з них для забезпечення безперебійної роботи утвореної ними LiFi мережі протягом заданого часу з визначеною замовником ймовірністю безвідмовної роботи.

СПИСОК ЛІТЕРАТУРИ

1. Badeel R., Subramaniam S. K., Hanapi Z. M., Muhammed A. A Review on LiFi Network Research: Open Issues, Applications and Future Directions. *Applied Sciences*. 2021. Vol. 11, no. 23, article no. 11118. P.1–35. DOI: 10.3390/app112311118.
2. Muraleedharan N., Cohen D. S. Modelling and simulation of UAV systems. *Imaging and Sensing for Unmanned Aircraft Systems: Control and Performance*. 2020. P. 101–121. DOI:10.1049/PBCE120F_ch5.
3. Udvardy P., Beszedes B., Toth B., Foldi A., Botos A. Simulation of obstacle avoidance of an UAV. *New Trends in Aviation Development (NTinAD'2020)* : Proc. 15th IEEE Int. Conf., 2020. P. 245–249. DOI: 10.1109/NTAD51447.2020.9379113.
4. Romaniuk S., Gosiewski Z., Ambroziak L. A ground control station for the UAV flight simulator. *Acta Mechanica et Automatica*. 2016. Vol 10, no. 1. P. 28–32. DOI: 10.1515/ama-2016-0005.
5. Zhenxiong W., Kai K., Xueying H., Huanming H., Jianghai L., Lang C. Computational simulation study on disturbance of six-rotor UAVs due to ammunition launch. *Journal of Physics: Conference Series*. 2023. Vol. 2478, article no. 102022. P. 1–19. DOI: 10.1088/1742-6596/2478/10/102022.
6. Chandrasekaran K., Theningaledathil V., Hebbar A. Ground based variable stability flight simulator. *Aviation*. 2021. Vol. 25, no. 1. P. 22–34. DOI: 10.3846/aviation.2021.13564.
7. Kampf R., Soviar J., Bartuška L., Kubina M. Creation of SW for Controlling Unmanned Aerial Systems. *LOGI - Scientific Journal on Transport and Logistics*. 2022. Vol. 13, no. 1. P. 198–209. DOI: 10.2478/logi-2022-0018.
8. Phadke A., Medrano F. A., Sekharan C. N., Chu T. Designing UAV Swarm Experiments: A Simulator Selection and Experiment Design Process. *Sensors*. 2023. Vol. 23, no. 17, article no. 7359. P. 1–26. DOI: 10.3390/s23177359.
9. Al-Mousa A., Sababha B. H., Al-Madi N., Barghouthi A., Younisse R. UTSim: A framework and simulator for UAV air traffic integration, control, and communication. *IJARS*. 2019. Vol. 16, no. 5. P. 1–19. DOI: 10.1177/1729881419870937.
10. Leichenko K., Fesenko H., Kharchenko V. Deploying the Reliable UAV Swarm for Providing P2P LiFi Communications Considering Physical Obstacles: Method of Rectangles, Algorithms, and Tool. *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2023)* : Proc. 12th IEEE Int. Conf., Dortmund, Germany, Sept. 07–09, 2023. P. 1011–1016. DOI: 0.1109/IDAACS58523.2023.10348819.
11. Leichenko K., Fesenko H., Borges J. A., Kharchenko V. Search for the Shortest Route Considering Physical Obstacles: Method of Controlled Waterfall, Tool, and Application. *Dependable Systems, Services and Technologies (DESSERT'2023)* : Proc. 13th IEEE Int. Conf., Athens, Greece, Oct. 13–15, 2023. In print.

Received (Надійшла) 11.11.2023
Accepted for publication (Прийнята до друку) 07.02.2024

A software tool to support the planning of UAV-based LiFi network deployment to ensure data transmission in the conditions of destruction

Kyrylo Leichenko, Herman Fesenko

Abstract. Accidents at critical infrastructure facilities are accompanied by damage to regular data transmission networks from sensors monitoring critical parameters of technological equipment to crisis centres. The absence of such data can lead to erroneous and insufficiently informed decisions by the crisis centre staff during actions to localise and eliminate the consequences of the accident. As an alternative to damaged regular networks, a LiFi network based on unmanned aerial vehicles (UAVs) can be considered, where the latter act as repeaters. However, due to the destruction of equipment and structures in production facilities, mechanical interference may occur, which will require the construction of LiFi signal propagation routes to bypass these obstacles. The subject of the article is the means of planning the deployment of flying networks to ensure data transmission in the conditions of destruction. The purpose of the article is to propose a software tool to support the planning of UAV-based LiFi network deployment to ensure data transmission in the conditions of destruction of critical infrastructure facilities. Objectives of the article: to propose a scheme for deploying a UAV-based LiFi network in a production facility with obstacles; to present the architecture and features of the software tool for implementing the proposed scheme; to provide examples of using the software tool. The following results were obtained. A software tool has been developed to support the planning of the deployment of a UAV-based LiFi network in industrial premises with obstacles, which allows you to lay routes for the propagation of LiFi signal (light flux with data) and determine the required number of UAVs and their location on the route. The results of applying a software tool for planning the deployment of a LiFi network in a given production facility with obstacles using the rectangle and controlled waterfall methods to bypass them and build routes are presented. The direction of further research is to develop a method and software tool for determining the required number of shifts and the number of UAVs in each of them to ensure the uninterrupted operation of the LiFi network formed by them for a given time with the customer-defined probability of the failure-free operation.

Keywords: unmanned aerial vehicle, LiFi network, software tool, obstacle avoidance, route planning.

UDC 621.9 **doi: 10.26906/SUNZ.2024.1.201**

Aleksandr Serkov, Vitalii Breslavets, Juliya Breslavets, Igor Yakovenko, Irina Yatsenko

National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

INFLUENCE OF PULSE ELECTROMAGNETIC RADIATION ON PERFORMANCE OF ELECTRIC RADIO PRODUCTS

Abstract. The **subject matter** is the processes of analysis of the occurrence of reversible and irreversible failures of semiconductor devices under conditions of exposure to electromagnetic radiation. It is shown that the influence of pulsed electromagnetic radiation is accompanied by the emergence of currents in the conductive elements of products and the emergence of their internal fields. The mechanisms for the occurrence of instabilities of natural oscillations of semiconductor components of electrical and radio products, caused by their interaction with flows of charged particles, have been determined. The presence of instabilities of this kind has a significant impact on the spectral (operating) characteristics of electrical radio products. The results obtained in the work make it possible to assess the degree of influence of pulsed electromagnetic radiation on the operating (volt-ampere) characteristics of electrical radio products. The **aim** is a model of the mechanisms of the emergence and development of instabilities of natural oscillations of semiconductor structures, components of electrical radio products (communication equipment), in the presence of currents and voltages induced by pulsed electromagnetic radiation. The implementation of such a model is due to the possibility of transforming the energy of a flow of charged particles induced by external electromagnetic radiation into the energy of natural vibrations of a semiconductor structure, taking into account the properties of the structure itself (size, concentration of free carriers, permeability). The transformation of the energy of currents induced by electromagnetic radiation into natural vibrations of a semiconductor structure is determined by two effects (transition or Cherenkov radiation) depending on the location of the structure relative to the direction of the currents. The **objectives** are: the main electromagnetic effects affecting performance of electrical radio products (ERI) under exposure conditions external pulsed radiation and also indicates characteristic changes ERI parameters that determine their functional purpose, which are a consequence of these effects. The **methods** used are the method of successive approximations over a small parameter, which allows one to determine the spectrum of natural oscillations of a semiconductor device and the mode of their amplification (instability). The following **results** are obtained. The results of studies characterizing the malfunction of electrical radio products under conditions of exposure to third-party electromagnetic radiation are presented, as well as the main parameters characterizing the electromagnetic resistance of electronic devices to the effects of pulsed currents and voltages. The characteristic types of malfunctions of semiconductor devices (SCD), components of electronic components, in areas of irreversible and reversible failures, as well as the levels of intensities and currents of electric and magnetic fields affecting the SPD, separating the areas of reversible and irreversible failures, are given. Using the energy approach, a physical model of the occurrence of one of the types of reversible failures of the semiconductor element base (the appearance of S-shaped sections of current-voltage characteristics) has been developed. This physical model makes it possible to determine the criteria for the electromagnetic resistance of a number of semiconductor devices to the effects of external pulsed radiation and also to obtain calculated ratios for assessing the degree of deviation of the operating characteristics of the PPP from the norm. **Conclusion**. Development of design relationships that determine the modes of amplification (generation) of oscillations of electrical radio products, making it possible to determine the degree of distortion of their current-voltage characteristics (reversible failures) and complete loss of performance (irreversible failures) depending on the parameters of external electromagnetic radiation. The results obtained in the work can be used in the development of amplifiers, generators and frequency converters operating in the millimeter and submillimeter range that are resistant to external electromagnetic radiation. Quantitative estimates of the criterion for reversible failures (instability increments) show that the amount of radiation energy lies within the sensitivity of modern submillimeter radiation receivers and is the cause of failures.

Keywords: semiconductor components, induced current, electromagnetic radiation, instability of oscillations, surface vibrations.

Introduction

All types of failures of radio-electronic equipment associated with the influence of third-party factors are usually divided into reversible and irreversible. Most of the available theoretical and experimental results of studies of the influence of electromagnetic radiation (EMR) on radio products belong to the field of irreversible failures, which are characterized by a complete loss of performance

At the same time, the development of interaction mechanisms between currents and voltages induced by electromagnetic radiation and processes characterizing the functional purpose of products is usually carried out within the framework of the theory of circuits with distributed parameters. This approach makes it possible to evaluate the performance criteria as a whole (for example, to estimate the critical energy characterizing thermal breakdown), however, questions related to the determination of various types of electromagnetic interactions occurring directly in the components of the product when exposed to EMR remain open.

The expansion of areas of application and increasing speed of electronic equipment (REE) leads to the need for an increasing use of the element base containing semiconductor electronics products [1]. The increase in the dependence of the performance of component equipment on the influence of electromagnetic radiation is associated with an increase in the complexity of the tasks that are assigned to electrical radio products, which leads to an increase in their sensitivity [1] This increases the degree of influence of external electromagnetic radiation on the performance of electronic devices, to the effects of

which semiconductor components have increased sensitivity. At the same time, the probability of reversible failures increases, which are characterized by a temporary loss of performance associated with a distortion of the output characteristics of the device.

One of the reasons for the occurrence of reversible equipment failures is caused by a change in the current-voltage characteristics of semiconductor devices - the appearance of areas with negative resistance in the forward current sections. The appearance of such areas is associated with the establishment of a mode of amplification of natural oscillations of the semiconductor structure - the possibility of transforming the energy of a flow of charged particles induced by external electromagnetic radiation into the energy of natural oscillations of the semiconductor structure. The increase in oscillations is characterized by an exponential increase in the amplitude of oscillations, i.e. their instability [3].

This work to a certain extent compensates for the existing gap in the research of reversible failures of this kind. It examines the interaction of flows of charged particles induced by pulsed electromagnetic radiation with wave processes in semiconductor structures used in modern radio electronics.

Task solution

The area of irreversible failures of semiconductor components of electrical and radio products has been studied in quite detail, both experimentally and theoretically [1]. The currently used methods for assessing the durability criteria of semiconductor devices in this failure region when exposed to external pulsed electromagnetic radiation (EMR) [2-3] consider in this capacity the value of the critical energy $W_{кр}$, exceeding which leads to a complete loss of performance (energy characterizing thermal breakdown. At the same time, to calculate currents and voltages induced by EMR, $W_{кр}$ which determine and arise directly in semiconductor components inside electrical and radio products, a transmission line model (generalized telegraph equations) is traditionally used. In this case, the components of electrical radio products are considered as linear circuits containing R, L, C - elements and include sources of currents and voltages caused by external electromagnetic influence. The circuit equations are solved in the frequency domain, and the inverse Fourier transform is used to move to the time domain [1]. It should be noted that semiconductor devices (SPDs) and integrated circuits (ICs) are most sensitive to the influence of external electromagnetic fields. Therefore, the main influence on the performance of this kind of electrical and radio products is exerted by voltages induced by external radiation. Irreversible failures of these products are usually associated with electrical (the magnitude and distribution of currents in the device structure, leading to breakdown) and thermal (an increase in the temperature of individual sections of the structure up to) [2].

This approach has a number of significant limitations and does not take into account effects that can affect the performance of semiconductor devices in the field of reversible failures, which does not allow reliably assessing the criteria for electromagnetic resistance [2]. These limitations are due to the fact that when analyzing reversible (short-term) failures in the performance of semiconductor devices, it is necessary to take into account the following factors of third-party electromagnetic influence [8]:

- dimensions (dimensions) of semiconductor components of ERI;
- time characteristics of the electromagnetic pulse (duration pulse rise and fall);
- constructive arrangement of installation of circuits relative to the shielded housing of the electrical radio product;
- orientation of the components of the semiconductor structure and vectors of the electric and magnetic pulse field;

It should be noted that the characteristic types of malfunctions of semiconductor devices (SCD), components of electrical and radio products, they depend on the levels of electric and magnetic fields of the applied pulse. So for the fields $E < 100$ *кВ/м*; $H < 600$ *А/м* – as a rule, reversible failures occur; otherwise, irreversible failures occur.

When determining electromagnetic immunity criteria, these factors cannot be fully taken into account within the framework of transmission line theory.

In particular, [1] the theory of long lines is limited to the low frequency range - the sizes of the structures under study are an order of magnitude smaller than the wavelengths. This limitation makes it impossible to describe the processes of interaction of radiation-induced currents with physical processes occurring directly in semiconductor components and determining their performance under conditions where the dimensions of the structures are comparable to the wavelengths.

In addition, transmission line theory generally assumes that all applied pulse energy is released at the critical circuit element, and the load is matched to the line. However, as a rule, the load is not matched over the entire frequency range and the low-power critical element is not located directly at the input, but after several passive elements capable of absorbing part of the radiation energy. Therefore, to determine the criteria for the occurrence of reversible failures (temporary loss of performance) within the framework of the theory of transmission lines, it is necessary to calculate the transient process for each specific device circuit.

Finally, the approach used (one-dimensional approximation) does not take into account the effects associated with the spatial limitation of semiconductor products and their location in relation to the direction of the applied pulse field (the orientation of the electromagnetic field and the direction of the vectors of operating currents and voltages in the components of the radio product themselves).

When considering the physical mechanisms of the occurrence of reversible failures of semiconductor components of radio products, this work proposes a more rigorous methodology based on the use of a complete system of electrodynamics (Maxwell)

equations, supplemented by material equations for the media that compose semiconductor devices, and boundary conditions that make it possible to determine the relationship between the quantities induced by external pulse of currents with the own electromagnetic fields of semiconductor devices. This approach makes it possible to study the processes of interaction of electromagnetic oscillations and induced currents directly in semiconductor components, which are not possible to describe within the framework of transmission line theory.

Thus, the object of research in this work was not considered circuits with lumped parameters, but limited conducting (semiconducting) media from which the components of electrical radio products are composed.

The results of this work to a certain extent compensate for the existing gaps in the research of one of the types of reversible failures - distortion of the volt-ampere characteristics of semiconductor devices under the influence of third-party electromagnetic radiation.

The reason for the occurrence of such distortions in the operating characteristics of devices is, in our opinion, the possibility of transforming the energy of currents induced by external EMR into the energy of the own electromagnetic oscillations of radio product components. In this case, the induced currents lose their energy by emitting natural oscillations of the semiconductor structure, and an increase in the forward current is accompanied by a voltage drop. Thus, an area with negative resistance appears on the current-voltage characteristic.

Surface waves existing at the boundaries of conducting solids were considered as a channel for transmitting the energy of currents induced by external radiation to semiconductor components.

The choice of surface oscillations as a channel for transmitting the energy of external radiation is not accidental - this type of waves is localized near the interfaces between media that compose semiconductor devices, so they transfer the energy of external electromagnetic fields more efficiently than volumetric oscillations [9].

The radiation mode (oscillation generation) in semiconductor devices usually manifests itself in the section of its current-voltage characteristic (volt-ampere characteristic), having a negative differential resistance (NDR) ($R = \frac{\Delta U}{\Delta I} < 0$) . In this case, the induced current (electron flow induced EMR in a semiconductor device) loses part of its energy ($W_{EL} < 0$); while the current increases ΔI accompanied by a voltage drop ΔU . The appearance of such deviations in the current-voltage characteristics characterizes one of the possible mechanisms of reversible failures.

Therefore, as an energy criterion for assessing the electromagnetic resistance of semiconductor devices in this region of reversible failures, the value of the emission energy of natural surface oscillations of semiconductor devices, caused by their interaction with induced external EMR currents, was considered $W_{rad} = W_{EL}$

$$W_{rad} = I_I U_I \Delta t_{ei} . \quad (1)$$

Thus, the value W_{rad} determining the degree of deviation of the current-voltage characteristic, is a quantitative characteristic of this type of reversible failure.

Where $I_I; U_I$ - induced current and voltage, respectively, Δt_{ei} - time of effective interaction of induced currents and natural oscillations of the semiconductor structure.

In this work, we investigated two possible mechanisms for converting the energy of moving charges (EMR-induced currents) into the energy of surface plasmons of semiconductor components of radio products - the effects of Cherenkov and transition radiation (the corresponding mutual configuration of the vectors of the strengths of the acting electric field and the direct current of the semiconductor device (diode) are shown in Fig. 1.)

The first mechanism for transforming the energy of moving charges into vibration energy, considered in the work, is the Vavilov-Cherenkov radiation effect. (Cherenkov radiation) [5].

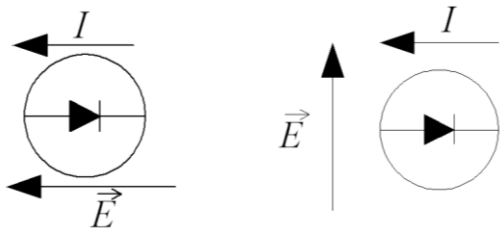
It is realized when induced currents move along the boundary of the semiconductor structure, and the phase velocity of the surface wave is equal to the velocity of charged particles (Fig. 1.2. - a). Under conditions of such resonance, the energy of induced currents (flow of charged particles) is transformed into the energy of natural oscillations of the ERE components and the oscillation generation mode is also established in the semiconductor device.

Another mechanism (transition radiation) is realized when the induced current (direction of the electric field strength vector of the acting external pulse) is perpendicular to the interfaces of the solid structure (semiconductor device) (Fig. 1.-b) and consists of the following [10].

When a charge moves in a material medium, the electromagnetic field it creates is determined not only by the magnitude of the charge and its speed, but also by the dielectric properties of the medium. If these properties change when a charge crosses the interface between media (semiconductor structure) at a constant speed, then the field created by the charge changes, part of the field is torn off from the particle and can be radiated into space. The resulting radiation is called transition radiation. As a result, when a stream of particles induced by an external electromagnetic field passes through a semiconductor structure, a continuous process of converting the energy of charges into the energy of natural oscillations of the field occurs. i.e., the oscillation generation mode is established in the structure.

It should be noted that recent experimental studies of transition radiation serve as the basis for the development of new methods for diagnosing flows of charged particles with high energy and, in addition, for solving problems of generation and amplification of electromagnetic oscillations [9]. The effect of transition radiation determines the mechanisms of excitation of a

wide variety of modes of semiconductor structures, so it becomes possible to transfer the energy of surface vibrations across the boundary, for which it is opaque in the absence of induced currents (i.e., exposure to external radiation).



a – Cherenkovsky radiation b – Transitional radiation

Fig. 1. Location of the applied electric field strength $\vec{E}$ and relative to forward current $\vec{I}$ semiconductor device (diode)

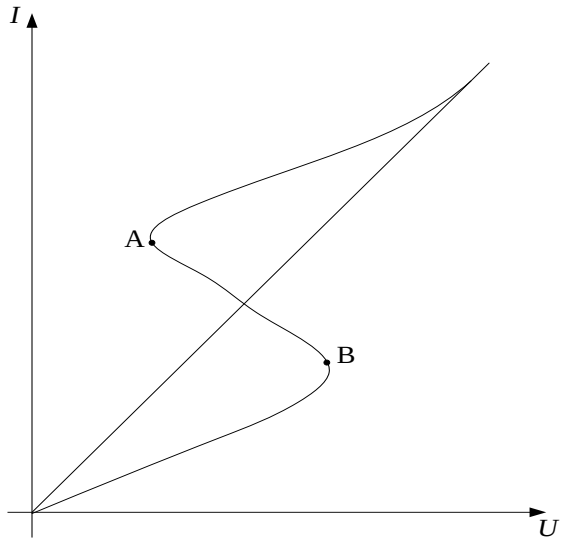


Fig. 2. Current-voltage characteristics of a semiconductor diode: CVC in the absence of external EMR; CVC in the presence of external EMR (reversible failure) (section AB)

Analysis

This mechanism of reversible failures (the appearance S - shaped volt-ampere characteristics of the direct current of the device (Fig. 2.) is realized in conditions where the amplitude strengths of the electric voltages E and magnetic H fields affecting the semiconductor radiation structure are in the range $E < 100\frac{kV}{m}$; $H < 600\frac{A}{m}$ [2].

In the case when the amplitudes of the strengths of the acting strength fields exceed the specified limits, the development of irreversible failures of the device is observed (thermal breakdown followed by melting and burnout of metallization and contact tracks [9].

The authors [7–8] solved a number of problems of interaction between waves and EMR-induced currents in semiconductor structures, which makes it possible to quantify the induced currents and voltages, as well as the time of effective interaction of the induced currents with the intrinsic fields of semiconductor structures. These quantities characterizing the radiation energy (the degree of deviation of the current-voltage characteristic from the norm) were determined within the framework of the theory of beam instabilities, since the oscillation generation mode is characterized by an exponential increase in the amplitude of the electro-magnetic fields of radiation from semiconductor devices:

$$E \approx \exp(+\gamma t);\ \gamma = 1/\Delta t_{ei}\ .$$

Here γ - instability increment, its value is determined by the parameters of the induced currents and the semiconductor device (concentration of current carriers and their speed).

Thus, solving problems of the emergence and development of beam instabilities in semiconductor structures (determining the growth rate of instabilities) allows you to build a physical model of the occurrence of one of the types of reversible failures of semiconductor devices under the influence of third-party electromagnetic radiation.

Using this model in [6-11], quantitative characteristics of the electromagnetic resistance of radio products were obtained (radiation energy value W_{rad}) in the field of reversible failures. Under conditions where the amplitude of the external pulsed electric field strength $E_0 \approx 15\ kV/m$, pulse duration $\Delta t_{imp} \approx 500ns$, then the magnitude of the radiation energy of natural vibrations of solid layered structures - ΔW_{rad} amounts to $\approx (10^{-7} - 10^{-8})wt$, those. with the sensitivity of modern microwave radiation receivers [7] ($10^{-10}\,wt$) is quite detectable and is the cause of reversible failures.

Conclusion

The main types of failures of electrical and radio products under the influence of pulsed electromagnetic radiation (reversible and irreversible failures) are given.

The results of studies characterizing the malfunction of electrical radio products under the influence of third-party electromagnetic radiation are presented, as well as the main parameters characterizing electromagnetic resistance to the effects of pulsed currents and voltages.

The main electromagnetic effects that affect the performance of radio products under external influence are presented, and characteristic changes in the parameters of radio products that determine their functional purpose, which are a consequence of the occurrence of these effects, are indicated.

A physical model of the occurrence of one of the types of reversible failures of semiconductor components of radio products (the appearance of S-shaped sections of current-voltage characteristics) is substantiated. It is based on the process of converting the energy of currents induced by external electromagnetic radiation into the energy of natural oscillations of the semiconductor structure (establishing the mode of generating natural oscillations). This physical model makes it possible to determine the criteria for the electromagnetic resistance of a number of semiconductor devices to the effects of external pulsed radiation and also to obtain calculated relationships for assessing the degree of deviation of the operating characteristics of semiconductor devices from the norm.

REFERENCES

1. Potylitsyn A.P. Transition radiation and diffraction radiation. Sumilaries and differences // Nucl. Instrum. Methods Phys.Res. – 2018. - v. 145, - P. 67.
2. Rule D.W., Fiorto R.B., Kimura W.D. Noninterceptive beam dignostics based on diffraction radiation // A I P Conf.Proc. – 2019. – v.590. – P.510.
3. Fiorito R.B., Rule D.W. Diffraction radiation diagnostics for moderate to hight energy beam // Proc.of the 4. Int. Symp. On Radiation From Relativic Electrons. – 2016. – v.155. – P.67.
4. Mkrthyan A.R. Coherent diffraction radiation from an electron bunch // Nucl. Ins. Meth. Phys. Res. B. – 2018. – v.56., - P.69.
5. Aronov I.E., Beletskii N.N. Fundamental steps of group velocity for 2D surface polaritons in high magnetic field // Chechoslovak Jornal of Physics. –2016.- Vol.46(S5), -P.2473-2474.
6. Perez-Rodrigues F., and Yampolskii V.A. Hesteresis del campo acustico excitado electromagneticamente en una pelicula metalica // X11 Congreso National de la SMCSV. Programa. Cancun, Mexico. – 2018
7. Krowne C.M., Blakey P.A. On the existence of submillimiterwave negative conductance in n – gallium arsenide diodes // J. Appl. Phys. –2019. - t.62 №6 - P. 2257 - 2266.
8. Kravchenko V.I., Yakovenko I.V., Generation of electromagnetic oscillations of a semiconductor structure under conditions of external electromagnetic influence // Bulletin of NTU "KhPI" -2018. – No. 21. – pp. 161–169.
9. Kravchenko V.I., Yakovenko I.V. Charged flow control particles. Induced electromagnetic radiation, on waveguide characterristics of semiconductor components electrical and radio products // Bulletin of NTU "KhPI" – 2018. – No. 27. – P.83–89.
10. Kravchenko V.I., Yakovenko I.V., Study of external influence. electromagnetic influence on waveguide characteristics semiconductor superlattice // Bulletin of NTU "KhPI" – 2019. – No. 29. – P.89–96.
11. Kravchenko V.I., Yakovenko I.V. Surface attenuation vibrations of semiconductor structures of electrical radio products under conditions influence of external electromagnetic influence // Bulletin NTU "KhPI" – 2010. – No. 27. – P.96–103.
12. Kravchenko V.I., Yakovenko I.V. Kinetic mechanisms interaction of surface vibrations with electronic conductivities semiconductor structures under the influence of external electromagnetic radiation // Bulletin of NTU "KhPI" – 2020. – No. 30. – pp. 103–111.

Received (Надійшла) 11.11.2023
Accepted for publication (Прийнята до друку) 24.01.2024

Вплив імпульсного електромагнітного випромінювання на працездатність електрорадіовиробів

О. А. Серков, В. С. Бреславець, Ю. В. Бреславець, І. В. Яковенко, І. Л. Яценко

Анотація. Предметом досліджень є процес аналізу появи оборотних та незворотних відмов напівпровідникових приладів за умов впливу електромагнітного випромінювання. Показано, що вплив імпульсного електромагнітного випромінювання супроводжується виникненням струмів у провідних елементах виробів та виникненням внутрішніх полів. Визначено механізми виникнення нестійкостей власних коливань напівпровідникових комплектуючих електрорадіовиробів, які зумовлені їхньою взаємодією з потоками заряджених частинок. Наявність нестійкостей такого роду істотно впливає на спектральні (робочі) характеристики електрорадіовиробів. Отримані в роботі результати дозволяють оцінити ступінь впливу імпульсного електромагнітного випромінювання на робочі (вольт-амперні) характеристики електрорадіовиробу. **Метою статті** є розробка моделі механізмів виникнення та розвитку нестійкостей власних коливань напівпровідникових структур, що комплектують електрорадіовироби (апаратуру зв'язку), за наявності струмів та напруг, наведених імпульсним електромагнітним випромінюванням. Реалізація подібної моделі обумовлена можливістю трансформації енергії потоку заряджених частинок, наведеного зовнішнім електромагнітним випромінюванням, в енергію власних коливань напівпровідникової структури з урахуванням властивостей самої структури (розмірами, концентрацією вільних носіїв, проникністю). Трансформація енергії наведених електромагнітним випромінюванням струмів у власні коливання напівпровідникової структури визначається двома ефектами (перехідного чи черенківського випромінювань) залежно від розташування структури щодо напрями струмів. **Задачі дослідження**: основні електромагнітні ефекти, що впливають на працездатність електрорадіовиробів (ЕРВ) в умовах впливу зовнішнього імпульсного випромінювання, а також зазначені характерні зміни параметрів ЕРВ, що визначають їх функціональне призначення, які є наслідком даних ефектів. **Використовувався метод** послідовних наближень за малим параметром, що дозволяє визначити спектр власних коливань напівпровідникового приладу та режим їхнього посилення (нестійкості). The following **results** are obtained. Наведено результати досліджень, що характеризують порушення функціонування електрорадіовиробів за умов впливу стороннього електромагнітного випромінювання, а також основні параметри, що характеризують електромагнітну стійкість ЕРВ до впливу імпульсних струмів та напруг. Наведено характерні типи порушень функціонування напівпровідникових приладів (НПП), комплектуючих ЕРВ, в областях незворотних та оборотних відмов, а також рівні напруженостей та струмів, що впливають на НПП електричних та магнітних полів, що розділяють області оборотних та незворотних відмов. З використанням енергетичного підходу розроблено фізичну модель виникнення одного з видів оборотних відмов напівпровідникової елементної бази (поява S – образних ділянок вольт – амперних характеристик). Дана фізична модель дозволяє визначати критерії електромагнітної стійкості низки напівпровідникових приладів до впливу зовнішнього імпульсного випромінювання, а також отримувати розрахункові співвідношення для оцінки ступеня відхилення робочих характеристик НПП від норми. **Висновки.** Розробка розрахункових співвідношень, що визначають режими посилення (генерації) коливань електрорадіовиробів, що дозволяють визначати ступінь спотворення їх вольт - амперних характеристик (зворотні відмови) та повної втрати працездатності (незворотні відмови) залежно від параметрів зовнішнього електромагнітного випромінювання. Результати, отримані в роботі, можуть бути використані при розробці стійких до впливу зовнішнього електромагнітного випромінювання підсилювачів, генераторів та перетворювачів частоти, що працюють у міліметровому та субміліметровому діапазоні. Кількісні оцінки критерію оборотних відмов (інкрементів нестійкостей) показують, що величина енергії випромінювання лежить у межах чутливості сучасних приймачів випромінювання субміліметрового діапазону і є причиною відмов.

Keywords: напівпровідникові комплектуючі, наведений струм, електромагнітне випромінювання, нестійкість коливань, поверхневі коливання.

UDC 621. 396.2 **doi: 10.26906/SUNZ.2024.1.206**

I. M. Syvolovskyi[1], V. P. Lysechko[1], O. S. Zhuchenko[1], O. M. Komar [2], V. V. P astushenko[1]

[1] Ukrainian State University of Railway Transport, Kharkiv, Ukraine
[2] National Aviation University, Kyiv, Ukraine

ANALYSIS OF METHODS FOR ORGANIZING DISTRIBUTED TELECOMMUNICATION SYSTEMS USING THE PARADIGM OF EDGE COMPUTING

Abstract: The article analyzes modern architectures of edge and fog computing systems, including OpenFog, F2c2C (Cloudlet), MELINDA, and architectures based on SDN and NFV. Particular attention is given to the study of Fog Computing from the conceptual and programmatic points of view. The advantages and limitations of the studied architectures in the context of IoT application are determined. Opportunities for enhancing telecommunication systems and improving the quality of service through the use of appropriate architectures are identified. The necessity of taking into account the specific needs and features of each system when choosing the appropriate fog computing architecture is proved. The need and relevance of further development and improvement of these architectures for optimal use are substantiated.

Keywords: N-tier architecture, edge computing, distributed systems, fog computing, Internet of Things, OpenFog, video stream processing, SDN, NFV, telecommunication system.

Abbreviations

IoT is an Internet of Things;
F2c2C is a Fog-to-cloudlet-to-Cloud;
MELINDA is a Multilevel Information Distributed Processing Architecture;
MLT is a Measurement Level Task;
FLT is a Feature Level Task;
DLT is a Decision Level Task;
SDN is a Software-Defined Network;
NFV is a Network Functions Virtualization;
SDNFV is a Software-Defined NFV
QoS is a Quality of Service.
ASTP is an Adaptive Selection and Task Priority.
SuVMF is a Software-defined Unified Virtual Monitoring Function.

Problem statement

Modern telecommunication systems that process large amounts of data and require minimal latency face the need to implement specialized architectures that allow for optimal resource utilization, improve service quality, and reduce delays. In particular, there is a need to research and implement Fog Computing architectures that facilitate efficient operation and reduce data transmission delays. These architectures also allow for the specific needs of different IoT systems and applications, making them more adaptable and productive. The study was necessitated by the need to optimize telecommunication systems in response to modern requirements and the growing amount of data in the IoT field [1, 2].

Analysis of recent research and publications. Despite the relevance of the study of edge computing architectures, there is currently no systematization in this area, no comprehensive analysis and comparative evaluation of different architectures, although many authors, both domestic and foreign, have partially studied this issue [1–11]. Given the rapid development of this area and its potential for the introduction of new technologies, the availability of such a study is a scientific and practical need.

The purpose of the article is to analyze potential solutions for improving distributed telecommunication systems using IoT and edge computing, as well as to justify the need to select optimal solutions for specific challenges and needs.

Presentation of the main material

In recent years, the availability of cloud technologies has led to the widespread integration of Cloud Computing into server systems, which has radically changed the paradigm of infrastructure and computing environments in business and technology fields, including telecommunications. The popularization of this paradigm and its widespread adoption was primarily due to the virtually unlimited expansion of server system resources through virtualization of all components.

In terms of architectural solutions, in the context of cloud-based server systems, the popularity of cloud integration has led to specific patterns or architectural solutions being used to organize these systems. One of these patterns is a two-tier architecture, which involves dividing the server system into two tiers: frontend and backend. The frontend is responsible for processing user requests and interacting with them, while the backend performs data operations and computations. This two-tiered architecture is a common approach that provides a certain level of standardization and allows for effective separation of functional responsibilities and scalability of the system (Fig. 1).

An architecture such as the one shown in Figure 1 is acceptable in the context of conventional client-server applications which a user interacts. But, in the context of IoT, tasks with the following parameters may arise.

1. Network bandwidth. A large number of connected IoT devices that constantly generate data can create significant problems with cloud network bandwidth, lead to overload and reduce the quality of service.

2. Latency. A significant distance between the IoT device and the cloud server can create a delay in data exchange. This can become a critical problem in cases requiring quick reactions, such as security systems or medical devices.

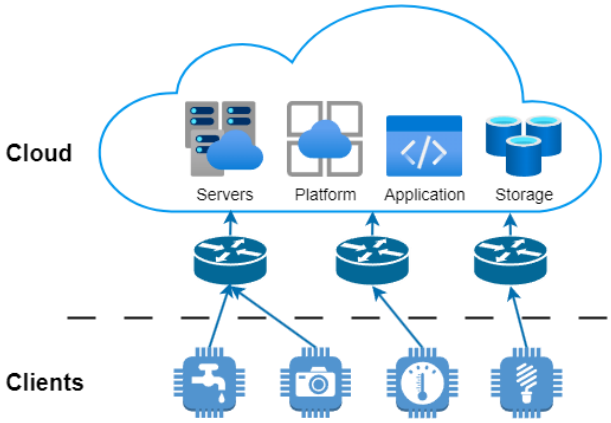


Fig. 1. Two-tier Cloud Computing architecture

3. Data security. The increased number of connected devices creates opportunities for cyberattacks and data breaches.

4. Scalability and administration. With a large number of connected devices, there are problems with administration and management.

5. Ensuring the viability of devices. A number of IoT devices have limited resources, such as batteries. Ensuring the longevity of the Internet of Things and their reliability requires the development of effective strategies for managing energy consumption and monitoring device health.

6. Interoperability and standardization. Different manufacturers may use different protocols to connect their devices to the IoT network, which can affect compatibility and integration between devices and systems.

To solve these problems, the Edge Computing paradigm was developed with the purpose of transferring part of the computing (functionality) to nodes that are closer to the devices than the cloud. At the same time, a computing node can be not only a data center, but any device with computing capabilities.

The emergence of Edge Computing set the general concept of such systems, which contributed to the development of new architectures later.

Fog Computing is an architecture concept in which an additional layer of processing nodes is added between the cloud layer and the device layer. It is often mentioned as a synonym for Edge Computing and is considered depending on the interpretation: both as an additional layer of the cloud layer and as an additional layer next to devices.

Dew Computing is a microservice concept that is embodied in a platform where devices can interact with each other continuously within a single "local" network, and this interaction takes place without sending data to cloud resources. Smart devices are one example of how this concept can be used.

Fog-Dew Computing is a synthesis of aforementioned architectures that utilizes its main advantages: devices operate as autonomous devices that do not require a constant connection to the Internet (to the cloud), but are connected to a local server. However, the local server interacts with cloud resources and is responsible for providing services to devices.

Nowadays, according to Google Scholar, Scopus, and Web of Science statistics, Fog Computing attracts the most attention from researchers due to its versatility and applicability in various areas of the Internet of Things (IoT) [3].

It should be noted that the concept of Dew Computing is limited, with opportunities for usage only in certain distributed telecommunication systems. Therefore, further research should focus on Fog Computing architectures.

When analyzing architectures, it is important to adhere to a clear definition of the concepts. It is essential to distinguish between the concepts of layer and tier, which are used synonymously in practice, but have a significant difference. A layer is a way of logically structuring the components of a software solution, while a tier is a way of physically structuring the infrastructure [4].

OpenFog N-tier architecture

In 2017, Princeton University and some leading IT companies from various industries formed the OpenFog consortium. The result of the collaboration was the OpenFog Reference Architecture document, which was the basis for the IEEE 1934-2018 protocol (Institute of Electrical and Electronics Engineers) [5].

This document is a new model of service architecture - FaaS (Fog as a Service), which includes the previously known ones: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), but with adaptation to Fog Computing, as well as ways to expand them.

In comparison to other architectures, OpenFog defines its advantages by using the SCALE acronym, which stands for.

1. Security - additional ways to achieve data security.

2. Cognition - ensuring autonomy by understanding the goals of the system's clients.

3. Agility - ensuring fast and affordable scaling.

4. Latency - reducing latency to ensure real-time processing.

5. Efficiency - dynamic allocation of system resources to achieve maximum efficiency.

The architecture is based on 8 basic principles called «pillars»: Security, Scalability, Openness, Autonomy, Programmability, RAS (Reliability, Availability, Serviceability), Agility, and Hierarchy. The description of each pillar is a set of recommendations and requirements for the system.

Fig. 2 demonstrates the OpenFog architecture with one level of Fog nodes.

According to Fig. 2, the OpenFog architecture does not impose any strict limitations on the number of layers. It makes possible to adapt the structure of the architecture to a specific subject area, so the number of cloud node tiers is arbitrary, and the presence of cloud and fog layers is optional.

Currently, despite the high level of standardization and description of this architecture, it is rarely used in practice due to the lack of a clear focus on a particular industry.

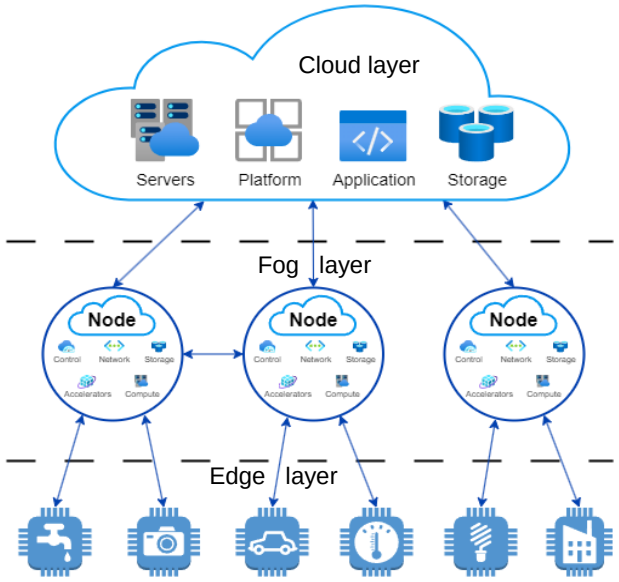


Fig. 2. N-tier OpenFog architecture

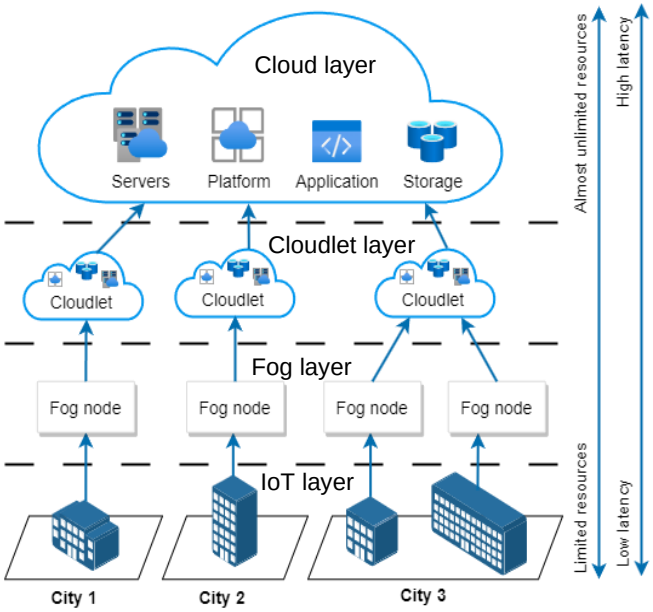


Fig. 3. Graphical representation of the F2c2C (Cloudlet) architecture on the example of the Smart City system

As a result, the main focus of current scientific and practical research is on Fog Computing architectures that are more adapted to specific domains and industries.

Cloudlets

The research of the OpenFog architecture has proven that one of the main challenges in the development of Fog Computing systems is to determine the optimal number of Fog node levels, their location, and allocation of resources for their operation.

The Smart City industry has traditionally used a centralized approach to organizing systems and data using Cloud Computing technologies. However, issues such as data protection, increased latency requirements, and energy efficiency have led to the consideration of decentralized architectures.

Some industries that are heavily utilizing the Internet of Things (IoT), primarily telecommunications systems, have the potential to develop distributed architectures at the conceptual level. For example, in the field of Smart City, researchers have introduced an innovation by proposing an architectural approach that offers territorial division (Fig. 3) [6].

Three conceptual levels are introduced to divide the system architecture by territorial basis: micro (building level), meso (neighborhood level), and macro (city or cloud level).

Also, in the context of this architecture, the concept of «cloudlet» is introduced, which implies a small data center that is located as close as possible to potential customers, unlike the cloud.

Since this architecture focuses on data management, three types of data are introduced according to their age.

1. Real-time data: produced by devices and nodes at the micro and meso levels in places where minimal latency is required.
2. Latest data: generated at the meso level, the result of real-time data processing.
3. Historical data: data stored in the cloud (macro level).

Guided by these concepts, one can define three main layers of this architecture (except for the device layer):

1. Fog layer: is as close to the devices as possible, works with real-time data, and located at the micro and meso levels.
2. Cloutlet-node layer: an intermediate layer between the cloud and fog, located in the same city as the devices (macro level), performs tasks of processing the «closest» data.
3. Cloud layer: processes and stores historical data, has unlimited resources.

Thus, this architecture combines the advantages of centralized and decentralized architectures: operation in a heterogeneous IoT environment, low load on the cloud network, the ability to process critical data in real time, etc.

MELINDA

One of the most complex systems in the telecommunications industry is video monitoring systems with real-time object detection. The traditional approach using Cloud Computing involves transferring the raw video stream to cloud data centers, where it is processed and then transmitted to the client. This approach has serious infrastructure-related drawbacks, such as.

1. High network saturation (each Full-HD camera generates a video stream of up to 12 Mbps), which creates problems with scaling the system in the form of limited cloud network bandwidth.
2. High and unstable latency when transferring data from the client to the cloud.
3. High resource and power consumption caused by the need to store a large amount of low-value data (video stream frames without objects or without changes).

To solve these problems, Neto A.R. proposed a three-tier Fog Computing architecture (Fig. 4) [7].

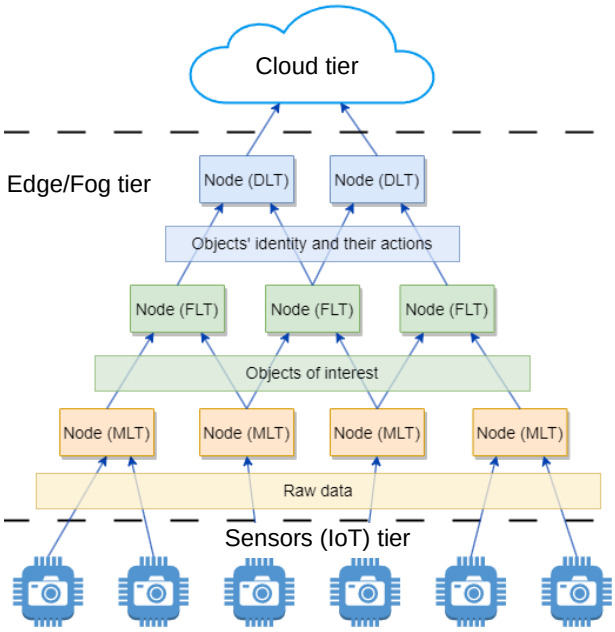


Fig. 4. Three-tier Fog Computing architecture optimized for video stream processing

According to Figure 4, the basic principle of this architecture is to reduce the amount of data by processing the video stream in multiple stages. Video stream processing consists of three steps.

– filtering the video stream to select only those frames that may contain an object;

– object identification;

– interpretation of the object's appearance (linking it to a specific event) for making decisions later.

Respectively, the architecture contains three types of processing nodes that perform tasks of different levels: Measurement Level Task (MLT), Feature Level Task (FLT), and Decision Level Task (DLT).

At the same time, this infrastructure is supported by the MELINDA (Multilevel Information Distributed Processing Architecture) software architecture, which consists of two subsystems.

1. Processing Subsystem, which consists of nodes for monitoring, resource allocation, and processing.

2. Management Subsystem, which consists of nodes for processing end-user requests (data extraction) and components for high-level system monitoring.

The Data Communication Manager component, common to both subsystems, is used for communication. It is assumed that each of the components is located on a separate node, and there can be several components of the same type.

SDN/NFV

The above architectures consider a Fog Computing system from a conceptual and programmatic point of view, following traditional methods of networking, using tree-like structures of Ethernet routers and mobile base stations. Many recent concepts for IoT include the use of new technologies: 5G as a data transmission technology, and SDN (Software Defined Network) and NFV (Network Functions Virtualization) for organizing network interaction [8].

The use of 5G is aimed at reducing the latency, cost, and power consumption of devices, as well as introducing new types of IoT systems that were previously impossible or difficult to implement.

Instead, the use of SDN and NFV enables us to look at some aspects of a Fog Computing system from a different perspective. Figure 5 shows the architecture of a system using Software-Defined NFV on a layer with Fog nodes.

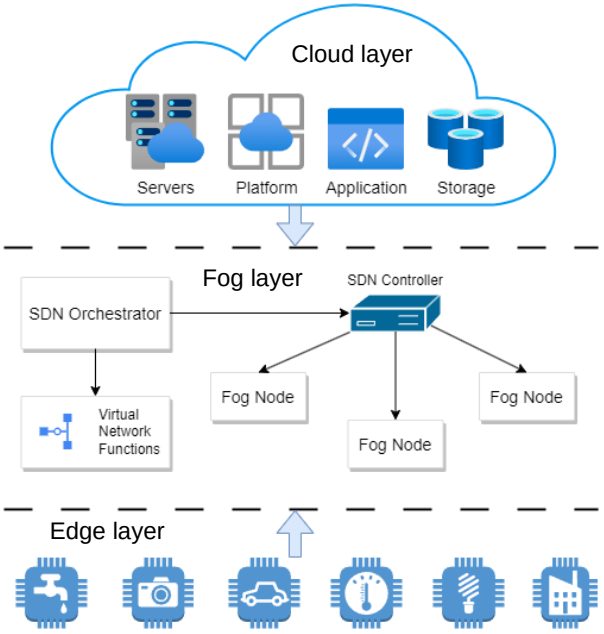


Fig. 5. System architecture using SDNFV on a layer with Fog nodes

Fig. 5 depicts that the basis of the fog layer is the SDN controller, which is responsible for processing all network traffic. It is assumed that this controller is an intermediate link between processing nodes and clients, and it is guided by 4 types of actions, namely: creating, modifying, executing, and terminating a task.

Thus, when a client requests a task from the orchestrator, it can check the validity of this request and request a certain amount of resources to execute this task.

In addition, it should be noted that the system takes into account five indices when executing tasks.

1. Cost index - the distance from the source to the node. The smaller the distance, the better the architecture.

2. Time index - the sum of the time spent on data transfer and the time of task execution. The smaller, the better.

3. Throughput index - the uniformity of user distribution in relation to traffic.

4. Energy Consumption index - the consumed energy, taking into account the energy consumption for system idle time. The lower, the better.

5. Capacity of machines in the cloud layer.

In order to test the performance of this architecture, a modeling environment consisting of MATLAB software and the EstiNet simulator was created. The main parameters characterizing the quality of the system were measured: total delay time, percentage of successfully completed tasks (reliability), and task processing speed (Quality of Service).

To analyze the performance of the proposed architecture, two existing similar architectures were selected, which are also based on SDN: ASTP (Adaptive Selection and Task Priority) and SuVMF (Software-defined Unified Virtual Monitoring Function) [9, 10].

The measurement results showed that the aforementioned architecture has higher values of the examined indicators compared to similar architectures: 90% reliability rate (vs. 85% in ASTP and 70% in SuVMF); 90% QoS (vs. 82% in ASTP and 68% in SuVMF). The results show that the use of SDN as a means of load balancing between resources can significantly improve system performance, which is highly relevant in various areas of IoT related to real-time processing, such as Industrial IoT [11]. Thus, further research on the use of this technology in IoT and fog computing systems is relevant.

Conclusion

The article investigates the methods of organizing distributed telecommunication systems using the Internet of Things (IoT), as well as the existing paradigms of Edge and Fog Computing.

During the analysis, the following fog computing architectures were reviewed: OpenFog, F2c2C (Cloudlet), MELINDA, and SDNFV architecture.

According to the results of the analysis, we can conclude that the researched architectures show the ability to solve the main problems of cloud architectures: high latency and high network saturation, by transferring part of the computation to nodes at the edge of the network.

In the field of distributed telecommunication systems, these architectures provide prospects for further development of performance improvement methods.

However, in practice, it is important to consider the specific requirements and features of each system when choosing a suitable fog computing architecture. In addition, the development and improvement of these architectures is an essential task, as they must be optimized for various applications, scenarios, and industries.

REFERENCES

1. Lysechko V., Zorina O., Sadovnykov B., Cherneva G., Pastushenko V.: Experimental study of optimized face recognition algorithms for resource – constrained. Academic journal: Mechanics Transport Communications, 2023, vol. 21, issue 1, article №2343, ISSN 2367-6620/
2. Lysechko V., Syvolovskyi I., Shevchenko B., Nikitska A., Cherneva G.: Research of modern NoSQL databases to simplify the process of their design. Academic journal: Mechanics Transport Communications, 2023, vol. 21, issue 2, article №2363, ISSN 2367-6620/
3. Naha R.K., Garg S., Georgekopolous D., Jayaraman P.P., Gao L., Xiang Y., Ranjan R.: Fog Computing: Survey of Trends, Architectures, Requirements, and Research Directions, IEEE Access, vol. 6, pp. 47980-48009, 2018.
4. Fowler M. Patterns of Enterprise Application Architecture. 1st Edition. – Addison-Wesley Professional, 2002. – 560 P. – ISBN 978-0321127426.
5. IEEE, “Ieee standard for adoption of openfog reference architecture for fog computing,” aug 2018, standard No. 1934-2018, Active since Aug. 2nd., 2018. [Online]. Available: https://standards.ieee.org/standard/1934-2018.html
6. Sinaeepourfard A., Krogstie J., Petersen S. A., Ahlers D.: F2c2C-DM: A Fog-to-cloudlet-to-Cloud, 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), pp. 590–595, 2019.
7. Neto A.R.: Edge-distributed Stream Processing for Video Analytics in Smart City Applications, 2021, 10.13140/RG.2.2.10968.57604.
8. Sreekanth G.R., Ahmed Najat Ahmed S., Sarac M., Strumberger I., Bacanin N., Zivkovic M.: Mobile Fog Computing by Using SDN/NFV on 5G Edge Nodes, Computer Systems Science and Engineering, vol. 41, pp. 751-765, 2021.
9. Wang J., Li D., Adaptive computing optimization in software-defined network-based industrial internet of things with fog computing, Sensors, vol. 18, no. 8, pp. 1–14, 2018.
10. Choi T., Kang S., Yoon S., Yang S., Song S. et al. «SuVMF: Software-defined unified virtual monitoring function for SDN-based large-scale networks» in Proc. CFI, ACM, Tokyo, Japan, pp. 1–6, 2014.
11. Alam M., Ahmed N., Matam R., Mukherjee M., Barbhuiya F.A.: SDN-Based Reconfigurable Edge Network Architecture for Industrial Internet of Things, IEEE Internet of Things Journal, 2023.

Received (Надійшла) 18.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

АНАЛІЗ МЕТОДІВ ОРГАНІЗАЦІЇ РОЗПОДІЛЕНИХ ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМ З ВИКОРИСТАННЯМ ПАРАДИГМИ ГРАНИЧНИХ ОБЧИСЛЕНЬ

І. М. Сиволовський, В. П. Лисечко, О. С. Жученко, О. М. Комар, В. В. Пастушенко

Анотація: В статті проаналізовано сучасні архітектури систем граничних та туманних обчислень, включаючи OpenFog, F2c2C (Cloudlet), MELINDA та архітектуру з використанням SDN та NFV. Особливу увагу приділено дослідженню Fog Computing з концептуальної і програмної точок зору. Визначено переваги та обмеження досліджених архітектур у контексті застосування в IoT. Виявлено можливості для вдосконалення телекомунікаційних систем і покращення якості обслуговування через використання відповідних архітектур. Доведено необхідність врахування конкретних потреб і особливостей кожної системи при виборі відповідної архітектури туманних обчислень. Обґрунтовано необхідність та актуальність подальшого розвитку та вдосконалення цих архітектур для оптимального використання.

Ключові слова: N-рівнева архітектура, граничні обчислення, розподілені системи, туманні обчислення, інтернет речей, OpenFog, Fog Computing, обробка відеопотоку, SDN, NFV, телекомунікаційна система.

UDC 621.39 **doi: 10.26906/SUNZ.2024.1.211**

Д. О. Шаманов, А. Р. Сорокін

Харківський національний університет радіоелектроніки, Харків, Україна

АНАЛІЗ СУЧАСНИХ МЕТОДІВ РАДІОЕЛЕКТРОННОЇ БОРОТЬБИ

Анотація. Метою даної роботи є аналіз сучасних методів радіоелектронної боротьби (РЕБ) та їх впливу на хід бойових дій. Актуальність дослідження зумовлена зростанням ролі РЕБ у сучасних війнах та конфліктах. У статті розглянуто такі ключові аспекти РЕБ, як російські методи РЕБ: аналіз основних розробок та їх використання у військовій доктрині, західні методи РЕБ: огляд основних розробок та їх застосування у військовій доктрині, та підбиття підсумків: порівняння російських та західних методів РЕБ, а також прогнози щодо розвитку РЕБ у майбутньому. Також проаналізовано особливості сучасних методів РЕБ, такі як широке використання мобільних систем, гнучкість застосування, інтеграція з іншими системами. Окремо розглянуто питання захисту від РЕБ.

Ключові слова: радіоелектронна боротьба, РЕБ, військова доктрина, мобільні системи, захист від РЕБ.

Вступ

В умовах стрімкої еволюції сучасних технологій та зростання загроз електронного характеру, радіоелектронна боротьба (РЕБ) стає важливим елементом військової стратегії та безпеки країни. У даній статті проводиться аналіз сучасних методів радіоелектронної боротьби.

Залежно від цілей і завдань, радіоелектронна боротьба може бути розділена на два основних види [1]:

- радіоелектронна розвідка (РЕР) - це комплекс заходів, спрямованих на отримання даних про радіоелектронні засоби (РЕЗ) противника. РЕР може проводитися як пасивно, шляхом виявлення і аналізу випромінювань РЕЗ противника, так і активно, шляхом постановки перешкод на шляху випромінювань;

- радіоелектронна протидія (РЕП) - це комплекс заходів, спрямованих на порушення або виведення з ладу РЕЗ противника.

РЕП може здійснюватися за допомогою різних засобів, таких як:

- радіоперешкоди - це створення в полі зору РЕЗ противника додаткового випромінювання, яке перешкоджає прийому корисного сигналу. Радіоперешкоди можуть бути генеровані як спеціальними засобами, так і за допомогою звичайних радіостанцій;

- радіозаваднення - це створення в полі зору РЕЗ противника сигналів, що імітують корисний сигнал. Радіозаваднення може використовуватися для дезорганізації роботи РЕЗ противника або для введення його в оману;

- радіопридушення - це нанесення фізичного ураження радіоелектронних засобів противника. Радіопридушення може здійснюватися за допомогою різних засобів, таких як лазерне або іонізуюче випромінювання.

Мета статті – ретельний огляд та аналіз сучасних методів радіоелектронної боротьби, включаючи ключові виклики та завдання цієї галузі.

Особлива увага приділяється російським науковим та військовим досягненням у сфері радіоелектронної боротьби, що дозволяє отримати глибше розуміння сучасного ландшафту електронного боротьби та його впливу на військові стратегії.

Історія розвитку та сучасний стан радіоелектронної боротьби

Радіоелектронна боротьба (РЕБ) - це комплекс заходів, спрямованих на вплив на радіоелектронні засоби противника.

РЕЗ - це будь-які засоби, що використовують електромагнітний спектр для передачі, прийому або обробки інформації.

Радіоелектронна боротьба (РЕБ) має довгу і бурхливу історію, яка почалася ще в Першій світовій війні.

У той час, перші радіостанції використовувалися для зв'язку між військами, і їхня робота була досить слабкою і легкою для перехоплення. Це призвело до розвитку перших засобів радіоперехоплення, які використовувалися для отримання інформації про противника.

Перший етап розвитку РЕБ (1914-1945 рр.) - це період зародження і розвитку РЕБ. На цьому етапі були розроблені перші засоби радіоперехоплення і радіопридушення, які використовувалися для отримання інформації про противника і для дезорганізації його роботи.

Другий етап розвитку РЕБ (1945-1991 рр.) - це період активного розвитку РЕБ. На цьому етапі були розроблені нові, більш ефективні засоби РЕБ, які були більш мобільними і могли використовуватися в різних умовах.

Третій етап розвитку РЕБ (з 1991 р. по сьогодення) - це період подальшого розвитку РЕБ. На цьому етапі використовуються нові технології, такі як штучний інтелект, квантові технології, тощо. Це дозволяє розробляти більш ефективні засоби РЕБ, які можуть використовуватися в різних видах бойових дій [1].

Огляд основних завдань радіоелектронної боротьби, розвідки та протидії

Основні завдання радіоелектронної боротьби можна розділити на два основних види:

- радіоелектронна розвідка (РЕР) - це комплекс заходів, спрямованих на отримання даних про РЕЗ противника. РЕР може проводитися як пасивно, шляхом виявлення і аналізу випромінювань РЕЗ противника, так і активно, шляхом постановки перешкод на шляху випромінювань;

- радіоелектронна протидія (РЕП) - це комплекс заходів, спрямованих на порушення або виведення з ладу РЕЗ противника.

Завдання РЕР включають в себе:

- виявлення, ідентифікацію та пеленгування РЕЗ противника - це дозволяє отримати інформацію про місцезнаходження, тип і характеристики РЕЗ противника;

- аналіз характеристик випромінювань РЕЗ противника - це дозволяє отримати інформацію про можливості та тактику застосування РЕЗ противника;

- отримання інформації про сигнали управління і даних РЕЗ противника - це дозволяє отримати інформацію про структуру управління і зв'язку противника, а також про його плани і цілі.

Завдання РЕП включають в себе:

- порушення роботи систем зв'язку і управління противника - це може ускладнити або унеможливити управління військами і бойовими діями противника, що може призвести до дезорганізації противника і його поразки;

- придушення роботи радарів противника - це може ускладнити або унеможливити використання авіації, протиповітряної оборони та інших систем противника, що може призвести до переваги союзників;

- пведення в оману противника - це може призвести до дезорганізації його дій або до введення його в помилкове рішення;

- фізичне ураження РЕЗ противника - це може призвести до їх повного знищення або виведення з ладу [2].

Вплив радіоелектронної боротьби на хід бойових дій

Радіоелектронна боротьба може мати значний вплив на хід бойових дій.

Вона може використовуватися для наступних завдань у ході бойових дій:

- порушення роботи систем зв'язку і управління противника - це може ускладнити або унеможливити управління військами і бойовими діями, що може призвести до дезорганізації противника і його поразки;

- придушення роботи радарів противника - це може ускладнити або унеможливити використання авіації, протиповітряної оборони та інших систем противника, що може призвести до переваги союзників;

- введення в оману противника - це може призвести до дезорганізації його дій або до введення його в помилкове рішення;

- фізичне ураження РЕЗ противника - це може призвести до їх повного знищення або виведення з ладу.

Загалом, РЕБ є складною і багатогранною областю. Вона постійно розвивається, і її ефективність постійно підвищується.

У сучасних умовах радіоелектронна боротьба є одним з найважливіших елементів військового мистецтва [3].

Огляд сучасних систем радіоелектронної боротьби

Українські розробки в галузі РЕБ мають давню історію. Ще в радянські часи в Україні існували потужні науково-дослідні інститути, які розробляли системи РЕБ для потреб Радянської Армії. Після розпаду СРСР Україна зберегла значний науковий та промисловий потенціал у галузі РЕБ. Українські підприємства продовжували розробляти та виробляти нові системи РЕБ, які відповідали сучасним вимогам.

Сьогодні Україна має на озброєнні широкий спектр систем РЕБ, які призначені для:

- придушення радарів противника;
- порушення роботи систем зв'язку противника;
- введення в оману РЕЗ противника;
- фізичного ураження РЕЗ противника.

Найвідоміші українські системи РЕБ:

- "Буковель" - комплекс РЕБ, призначений для придушення радарів противника;
- "Кольчуга" - комплекс РЕБ, призначений для постановки активних перешкод радарам противника (рис. 1);
- "Нота" - комплекс РЕБ, призначений для пеленгування та радіорозвідки;
- "Антей" - комплекс РЕБ, призначений для захисту літаків від радіолокаційних ракет.

Українські системи РЕБ активно використовувалися у війні на Донбасі та під час повномасштабної збройної агресії російської федерації. З їх допомогою українським військам вдалося:

- знизити ефективність роботи російських радарів;
- порушити роботу російських систем зв'язку;
- ввести в оману російські РЕЗ;
- захистити українські літаки від радіолокаційних ракет [4].

Рис. 1. Комплекс РЕБ «Кольчуга»

Україна продовжує розвивати свої системи РЕБ. Вже зараз українські розробники працюють над створенням нових систем РЕБ, які відповідатимуть сучасним вимогам. Важливо пам'ятати, що інформація про українські системи РЕБ є обмеже-

ною. Це пов'язано з тим, що багато систем РЕБ мають секретний характер.

Однак, навіть з наявної інформації можна зробити висновок, що Україна має потужний потенціал у галузі РЕБ.

Збройні сили США мають на озброєнні широкий спектр систем РЕБ, які призначені для:

- придушення радарів противника;
- порушення роботи систем зв'язку противника;
- введення в оману РЕЗ противника;
- фізичного ураження РЕЗ противника.

До складу американських систем РЕБ входять:

- стаціонарні системи РЕБ;
- мобільні системи РЕБ;
- авіаційні системи РЕБ;
- корабельні системи РЕБ.

Найвідоміші американські системи РЕБ:

- EA-18G Growler - літак РЕБ, призначений для придушення радарів противника;
- EC-130H Compass Call - літак РЕБ, призначений для постановки активних перешкод радарам противника (рис. 2);
- AN/ALQ-126 - система РЕБ, призначена для захисту літаків від радіолокаційних ракет;
- AN/SLQ-32(V) - система РЕБ, призначена для захисту кораблів від протикорабельних ракет.

Американські системи РЕБ показали свою високу ефективність у багатьох військових конфліктах.

Рис. 2. Літак РЕБ «EC-130H Compass Call»

Західні методи РЕБ характеризуються наступними особливостями:

- широке використання автоматизованих систем РЕБ. Автоматизовані системи РЕБ дозволяють швидко і точно виявляти, ідентифікувати та придушувати РЕЗ противника;
- інтеграція РЕБ з іншими системами. РЕБ інтегрується з іншими системами, такими як системи зв'язку, управління та озброєння. Це дозволяє підвищити ефективність застосування РЕБ;
- розвиток засобів захисту від РЕБ. Західні країни розробляють і впроваджують засоби захисту від РЕБ, які дозволяють зменшити вплив засобів РЕБ противника.

Західні країни є одними з провідних світових розробників засобів РЕБ [2]. Західні війська використовують широкий спектр засобів РЕБ, які включають в себе:

- радіоперехоплювальні станції, які дозволяють отримувати інформацію про РЕЗ противника;
- радіоперешкодні станції, які дозволяють придушувати роботу РЕЗ противника;
- радіозавадні станції, які дозволяють вводити в оману РЕЗ противника;
- системи радіопридушення, які дозволяють виводити з ладу РЕЗ противника.

Серед основних західних розробок у сфері РЕБ можна виділити наступні:

- система радіопридушення AN/ALQ-99, яка призначена для придушення роботи радарів противника, в тому числі радарів ППО (рис. 3);
- система радіоприглушення AN/ALQ-211, яка призначена для виведення з ладу РЕЗ противника, в тому числі радарів, систем зв'язку та управління.
- система радіоперехоплення AN/TSQ-171, яка дозволяє отримувати інформацію про РЕЗ противника, в тому числі про сигнали управління і даних [5].

Рис. 3. Система радіоприглушення AN/ALQ-99 встановлена на винищувачі F-16

Російські методи РЕБ характеризуються наступними особливостями [3]:

- широке використання мобільних систем РЕБ. Російські війська мають значну кількість мобільних систем РЕБ, які можуть швидко переміщатися по полю бою. Це дозволяє їм швидко реагувати на зміни обстановки і забезпечувати ефективну підтримку військ у ході бойових дій;
- гнучкість застосування РЕБ. Російські війська використовують РЕБ для вирішення широкого спектру завдань, включаючи придушення радарів, систем зв'язку та управління, а також введення в оману противника;
- інтеграція РЕБ з іншими системами. Російські війська інтегрують засоби РЕБ з іншими системами, такими як системи зв'язку, управління та озброєння. Це дозволяє підвищити ефективність застосування РЕБ.

Російська федерація є одним з провідних світових розробників засобів РЕБ. Російські війська використовують широкий спектр засобів РЕБ, які включають в себе:

- радіоперехоплювальні станції, які дозволяють отримувати інформацію про РЕЗ противника;
- радіоперешкодні станції, які дозволяють придушувати роботу РЕЗ противника;

- радіозавадні станції, які дозволяють вводити в оману РЕЗ противника;

- системи радіопридушення, які дозволяють виводити з ладу РЕЗ противника.

Серед основних російських розробок у сфері РЕБ можна виділити наступні:

- система радіопридушення "Красуха-4", яка призначена для придушення роботи радарів противника, в тому числі радарів ППО (рис. 4);

Рис.4. Система радіоелектронної боротьби «Красуха-4»

- система радіопридушення "Рись", яка призначена для придушення роботи систем зв'язку противника;

- система радіоприглушення "Биліна", яка призначена для виведення з ладу РЕЗ противника, в тому числі радарів, систем зв'язку та управління.

Висновки

У ході аналізу сучасних методів радіоелектронної боротьби виявлено, що ця галузь стає все важливішою у контексті сучасних військових конфліктів та загроз кібербезпеки. Застосування штучного інтелекту, кіберзаходів та новітніх технологій у РЕБ зумовлює необхідність постійного вдосконалення та адаптації стратегій електронного протистояння. Як російські, так і західні методи радіоелектронної боротьби відзначаються високим рівнем технічної складності та ефективності. Розгляд західних підходів свідчить про активну роль країн Заходу у розробці та впровадженні інноваційних технологій для контролю та протидії електронним загрозам. РЕБ є важливою складовою сучасного бою, і її значення буде лише зростати в майбутньому. Розвиток нових технологій дозволить створювати більш ефективні засоби РЕБ, які матимуть значний вплив на хід бойових дій.

СПИСОК ЛІТЕРАТУРИ

1. Палий А. И. Очерки истории радиоэлектронной борьбы. — Москва : Вузовская книга, 2006. — 284 с.
2. Добыкин В.Д., Куприянов А.И., Пономарёв В.Г., Шустов Л.Н. Радиоэлектронная борьба. Силовое поражение радиоэлектронных систем. — М. : Вузовская книга, 2007. — 468 с.
3. Пічугін М. Ф., Носова Г.Д. Збірник наукових праць ЖВІ НАУ. Випуск 3 — Аналіз тактики застосування підрозділів РЕБ у сучасних війнах та локальних збройних конфліктах. — К., 2010.
4. О. М. Семененко, Р. В. Бойко, Ю. Б. Добровольський, В. Л. Іванов, О. І. Кремешний. Контррадіоелектронна боротьба як складова частина радіоелектронної боротьби в Збройних Силах України // Системи озброєння і військова техніка. — 2016. — Вип. 46. — С. 141-145
5. И. А. Черепнев, Г. В. Фесенко, Г. А. Ляшенко, Н. В. Полянова, О. А. Макогон. Аналитический обзор состояния радиоразведки в начале XX века и боевого применения в первую мировую войну // Системи озброєння і військова техніка. — 2015. — Вип. 44. — С. 123-133.
6. Добыкин В. Д., Куприянов А. И., Пономарёв В. Г., Шустов Л. Н. Радиоэлектронная борьба. Силовое поражение радиоэлектронных систем. — М.: Вузовская книга, 2007. — 468 с.

Received (Надійшла) 21.11.2023
Accepted for publication (Прийнята до друку) 31.01.2024

Analysis of modern methods of electronic warfare

D. Shamanov, A. Sorokin

Abstract. The purpose of this paper is to analyze modern methods of electronic warfare (EW) and their impact on the course of hostilities. The relevance of the study is due to the growing role of electronic warfare in modern wars and conflicts. The article discusses such key aspects of electronic warfare as Russian electronic warfare methods: analysis of the main developments and their use in military doctrine, Western electronic warfare methods: review of the main developments and their use in military doctrine, and summarizing: comparison of Russian and Western electronic warfare methods, as well as forecasts for the development of electronic warfare in the future. The author also analyzes the features of modern electronic warfare methods, such as the widespread use of mobile systems, flexibility of use, and integration with other systems. Separately, the issue of protection against EW is considered.

Keywords: Electronic warfare, EW, military doctrine, mobile systems, EW protection.

АЛФАВІТНИЙ ПОКАЖЧИК

Наукове видання

СИСТЕМИ УПРАВЛІННЯ, НАВІГАЦІЇ ТА ЗВ′ЯЗКУ

Збірник наукових праць

Випуск 1 (75)

Відповідальна за випуск *К. С. Нестеренко*
Технічний редактор *Т. В. Уварова*
Коректор *О. В. Морозова*
Комп`ютерна верстка *Н. Г. Кучук, І. Ю. Петровська*
Оформлення обкладинки *І. В. Ільїна*

Свідоцтво про державну реєстрацію КВ № 19512-93/2ПР від 16.11.2012 р.

Підписано до друку 14.02.2024. Формат 60×84/8. Ум.-друк. арк. 27,0. Тираж 140 прим. Зам. 214-24
Адреса редакції: Україна, 36011, м. Полтава, Першотравневий проспект, 24, тел. (050) 302-20-71
Національний університет «Полтавська політехніка імені Юрія Кондратюка»

Віддруковано з готових оригінал-макетів у цифровій друкарні Impress
61002, м. Харків, вул. Пушкінська, 56, тел. **+ 38 (057) 714-52-11**
e-mail: **irina@impress.biz.ua**